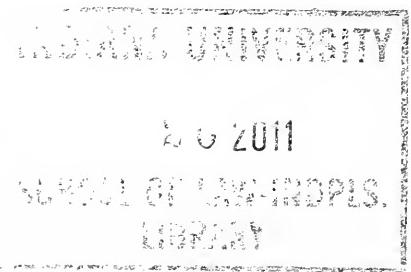

TRIBUTES

A Tribute to Justice Theodore Boehm
Gerald L. Bepko

Tribute to Justice Theodore R. Boehm
Joel M. Schumm
Paul L. Jefferson

Professor Henry C. Carlson
A Remembrance
October 29, 2010
William F. Harvey



ARTICLES

Judicial Strategy and Legal Reason
Evelyn Keyes

A Case of Doubtful Certainty: The Court Relapses into Search Incident to Arrest
Confusion in *Arizona v. Gant*
George M. Dery III

Towards Recognizing and Reconciling the Multiplicity of Values and Interests in Trademark Law
Michael S. Mireles, Jr.

Wrongfully Incarcerated, Randomly Compensated—How to Fund Wrongful-Conviction Compensation Statutes
Deborah Mostaghel

NOTES

Who Needs Contract Law?—A Critical Look at Contractual Indemnification (or Lack Thereof) in
FHAA and ADA “Design and Construct” Cases
Charles Daugherty

Application or Registration?: Confusion Regarding the Copyright Act’s Prerequisite to Copyright
Infringement Lawsuits
Thomas M. Landrigan

Justice or Mental Health . . . Should Litigants Have to Choose?
Mental Health as a Reason to Proceed Anonymously
Sarah Orme

“Growing Pains” in Indiana Age Discrimination Law
David T. Vlink



INDIANA UNIVERSITY
SCHOOL OF LAW-INDIANAPOLIS
IUPUI



BNA's Law School Program

**Multiple platforms, more choices
—for staff and students.**

Delivery of timely information is one of the cornerstones of BNA's success. Now, BNA delivers services electronically, through IP authentication, to all law school faculty staff and their students.

By subscribing to any of four law school-wide electronic packages or building your own custom package, you can give your students access to the same valuable information as practicing attorneys. They'll conduct research and follow changes in the law in the same way senior partners do. Your students will not only be prepared for internships and clerkships, they'll be better equipped for any job.

To learn more about BNA's law school program, call BNA Customer Service at **800-572-3374** or visit us at **www.bna.com/lawschool**.



BNA®

Essential information. Expert analysis.

Indiana Law Review

Volume 44

2010-2011

Editor-in-Chief

Kate Mercer-Lawson

Executive Managing Editor

Sara M. Benson

Executive Notes Editor

Sarah Orme

Senior Executive Editor

Jennifer Ekblaw

Executive Articles Editors

Charles B. Daugherty

Mac Schilling

Symposium Editor

Amanda L.B. Mulroony

Articles Editors

Lauren Berger

Zachary G. Jacob

Jared W. Josleyne

Jordan Szymialis

Perry Whan

Note Development Editors

Courtney L. Andreone

Thomas Landrigan

Kyle D. McClammer

Timothy J. Moriarty

Justin Spack

David T. Vlink

Associate Editors

Jonathan T. Faust

Brian K. Jarman

Members

Ted Batson, Jr.
Kory Bell
Courtney Benson-Kooy
Sean Brown
Amanda Dalton
Matthew Dinn
Erica Drew
Nick Dugan
Ian Fleming

Sarah Fowler
Matthew Gaudin
Molly Gilland
Jared Haller
John Higgins
Janelle Kilies
Lisa Laws Kohli
Lara Langeneckert
Michael Mullins
Keith Mundrick

Meghan Pitcher
Sean Roth
Kirk Severson
James Smerbeck
Brian Tekulve
Alicia Wanker
Maritza Webb
Alex Whitted
Beau Zoeller

Editorial Specialist

Chris Paynter

Faculty Advisors

Andrew R. Klein
R. George Wright

INDIANA LAW REVIEW

(ISSN 0090-4198)

Published four times a year by Indiana University. Editorial and Business Offices are located at:

Indiana Law Review
Lawrence W. Inlow Hall
530 W. New York Street
Indianapolis, IN 46202-3225
(317) 274-4440

Subscriptions. Current subscription rates for an academic year are \$30.00 (domestic mailing) and \$35.00 (foreign mailing) for four issues. Unless the Business Office receives notice to the contrary, all subscriptions will be renewed automatically. ***Address changes must be received at least one month prior to publication to ensure prompt delivery and must include old and new address and the proper zip code.***

Single and Back Issues. Current single regular issues may be obtained from the *Indiana Law Review* Business Office for \$10.00. Current Survey issues may be purchased for \$20.00. The Symposium Issue may be purchased for \$15.00. Please enclose payment with order. Claims for nonreceipt of current year's issues must be made within six months of the mailing date. Back issues may be ordered from: William S. Hein & Co., Inc., 1285 Main St., Buffalo, NY 14209-1987, (800) 828-7571, (716) 883-8100 fax, mail@wshein.com (e-mail for orders). Back issues are also available in an image-based format through HeinOnline, www.heinonline.org and Ebsco Publishing, www.ebscohost.com.

Manuscripts. The *Indiana Law Review* invites submission of unsolicited manuscripts. All citations should conform to *The Bluebook A Uniform System of Citation* (19th ed. 2010). Manuscripts should be double-spaced with standard one-inch margins. The *Indiana Law Review*, as a matter of policy, promotes the use of gender-neutral language, and alternate use of the female or male pronoun is intended to refer to both female and male.

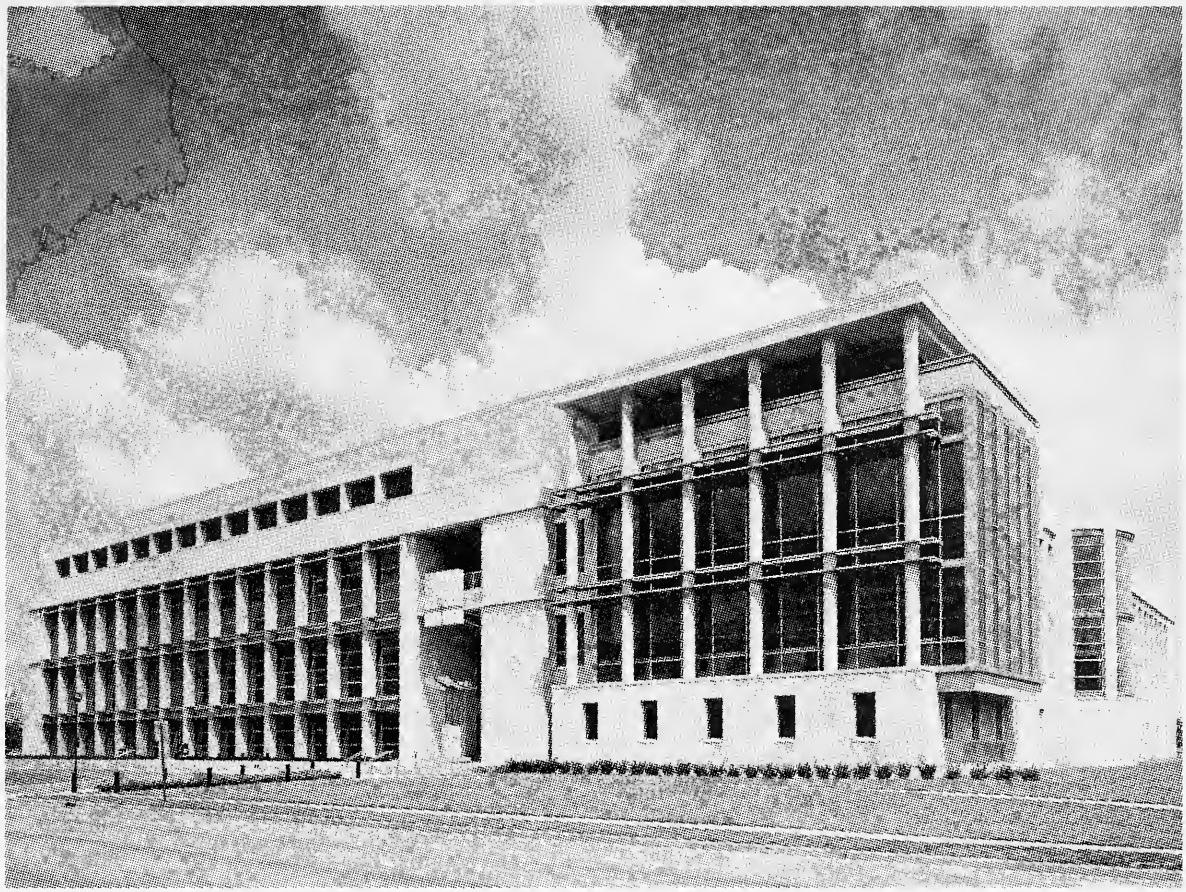
Copyright: Copyright 2011. The Trustees of Indiana University. Except as may be expressly provided elsewhere in this publication, permission is hereby granted to produce and distribute copies of individual works from this publication for nonprofit education purposes, provided that copies are distributed at or below cost, and that the author, source, and copyright notice are included on each copy. This permission is in addition to rights of reproduction granted under Sections 107, 108, and other provisions of the U.S. Copyright Act.

The INDIANA LAW REVIEW (ISSN 0090-4198) is the property of Indiana University and is published quarterly by the Indiana University School of Law—Indianapolis, which assumes complete editorial responsibility thereof. Current subscription rates for an academic year are: one year domestic \$30.00; foreign \$35.00. Send all correspondence to Editorial Specialist, *Indiana Law Review*, Indiana University School of Law—Indianapolis, Lawrence W. Inlow Hall, 530 W. New York Street, Indianapolis, Indiana 46202-3225. Publication office: 530 W. New York Street, Indianapolis, Indiana 46202-3225. Periodicals postage paid at Indianapolis, Indiana 46201.

POSTMASTER: Send address changes to INDIANA LAW REVIEW, Lawrence W. Inlow Hall, 530 W. New York Street, Indianapolis, Indiana 46202-3225.



The entire text of this Law Review is printed on recycled paper.



**Please enter my subscription
to the
*INDIANA LAW REVIEW***

Name _____

Address _____

Enclosed is \$_____ for _____ subscription(s)

Mail to: ATTN: Editorial Specialist
 at

INDIANA LAW REVIEW
INDIANA UNIVERSITY SCHOOL OF LAW—INDIANAPOLIS
Lawrence W. Inlow Hall
530 West New York Street
Indianapolis, Indiana 46202-3225

For an *academic* year, the subscription rate for four issues is:
Domestic, \$30; Foreign, \$35; Student, \$20

Single Issue, \$10; Survey Issue \$20
Symposium Issue, \$15

Digitized by the Internet Archive
in 2011 with funding from
LYRASIS Members and Sloan Foundation

Indiana International & Comparative Law Review

The *Indiana International & Comparative Law Review*, published by the Indiana University School of Law - Indianapolis since 1991, is a student-edited law journal that provides a forum for the discussion and analysis of contemporary issues in public and private international law. The *Review* publishes articles by prominent legal scholars, practitioners, and policy makers from around the world, as well as student-written notes and comments.

The *Review* is published twice annually, with a periodic third issue devoted to a specific topic in international or comparative law. In 1994, in conjunction with the Italian Academy for Advanced Studies in America at Columbia University, the *Review* published a special issue on the political and social aspects of Italian law. In 1996, the *Review* published an issue addressing various aspects of Chinese law.

Name _____

School/Firm/Business

Address

City, State, Zip Code

Subscriptions within the U.S. are \$18.00 per year and outside the U.S. are \$21.00 per year. I enclose \$ _____ for _____ subscription(s) to the *Indiana International & Comparative Law Review*. Please make your check payable to *Indiana International & Comparative Law Review*.

Mail to: Executive Production Editor
Indiana International & Comparative Law Review
Lawrence W. Inlow Hall
530 West New York Street
Indianapolis, Indiana 46202

In 2003, Indiana University School of Law - Indianapolis established a specialty law review focused specifically on health care law and policy:

INDIANA HEALTH LAW REVIEW

The scope of the INDIANA HEALTH LAW REVIEW includes bioethics, malpractice liability, managed care, anti-trust, health care organizations, medical-legal research, legal medicine, food and drug, and other current health-related topics.

Detach and mail this subscription form today to receive future issues

Subscription Price: \$20.00

- Two (2) Issues
- Automatic Renewal
- Foreign Mailing Is Extra

Mail form to:

Indiana Health Law Review
Indiana University School of Law – Indianapolis
Attn: Executive Business Editor
Lawrence W. Inlow Hall, Room 136C
530 W. New York Street
Indianapolis, IN 46202-3225

Questions?

Email: ihlr@iupui.edu
Phone: (317) 278-9519
www.indylaw.indiana.edu/~ihlr

Mailing Address:

Name/Title _____

Company/School _____

Address _____

City _____ State _____ ZIP _____

Email _____

Payment Method: Enclosed

(Make Check Payable to
INDIANA HEALTH LAW REVIEW)

Bill Me Later

Indiana University School of Law—Indianapolis

2010-2011 ADMINISTRATIVE OFFICERS AND FACULTY

Administrative Officers

- MICHAEL A. MCROBBIE, *President of the University*. Ph.D., Australian National University.
- CHARLES R. BANTZ, *Chancellor, Indiana University-Purdue University—Indianapolis*. B.A., M.A., University of Minnesota; Ph.D., The Ohio State University.
- GARY R. ROBERTS, *Dean and Gerald L. Bepko Professor of Law*. B.A., Bradley University; J.D., Stanford University.
- PAUL N. COX, *Vice Dean and Centennial Professor of Law*. B.S., Utah State University; J.D., University of Utah; LL.M., University of Virginia.
- JAMES P. NEHF, *Associate Dean for Graduate Studies and Cleon H. Foust Fellow; Professor of Law and John S. Grimes Fellow*. B.A., Knox College; J.D., University of North Carolina Law School.
- JONNA KANE MACDOUGALL, *Assistant Dean for External Affairs and Alumni Relations*. B.A., M.A., Indiana University; J.D., Indiana University School of Law—Indianapolis.
- JOHNNY PRYOR, *Assistant Dean for Student Affairs*. B.A., Wittenberg University; J.D., Indiana University Maurer School of Law.
- CHASTITY THOMPSON, *Assistant Dean for Professional Development*. A.S., B.A., Alabama State University; M.B.A., Auburn University; J.D., Indiana University School of Law—Indianapolis.
- MARK V. WUNDER, *Assistant Dean for Development*. B.S., J.D., University of Iowa.
- ELIZABETH ALLINGTON, *Director of Communications and Creative Services*. B.A., Indiana University; M.A., M. Phil., New York University.
- TERESA (TERRI) J. CUELLAR, *Director of Technology Services*. B.S., St. Bonaventure University, New York.
- JOYCE M. HERTKO, *Director of Major Gifts*. B.A., Knox College; M.S., Western Illinois University; Ph.D., Indiana University.
- AMANDA KAMMAN, *Director of Fund Raising and Development Services*. B.A., Indiana University.
- PRISCILLA D. KEITH, *Director of Research and Projects, Hall Center for Law and Health*. B.S., Spelman College; M.S., Atlanta University; J.D., Indiana University School of Law—Indianapolis.
- PATRICIA K. KINNEY, *Director of Admissions*. B.S., Purdue University; J.D., Indiana University School of Law—Indianapolis.
- VIRGINIA MARSCHAND, *Director of Administrative and Fiscal Affairs*. B.A., M.P.A., Indiana University; J.D., Indiana University School of Law—Indianapolis.
- SONJA RICE, *Egypt Program Director*. B.A., Purdue University; J.D., Indiana University School of Law—Indianapolis.
- JOHN R. SCHAILY, III, *Executive Director of The Center for Intellectual Property Law and Innovation*. B.A., Purdue University; J.D., Indiana University Maurer School of Law.
- LAWANDA W. WARD, *Director of Pro Bono Program and Public Interest*. B.A., Murray State University; M.A., Illinois State University; M.S., Old Dominion University; J.D., Indiana University School of Law—Indianapolis.
- JACOB MANALOOR, *Associate Director for Contracts, Grants and Fund Raising*. B.S., Indiana University Kelley School of Business; J.D., Indiana University School of Law—Indianapolis.
- CARLOTA TOLEDO, *Associate Director of Student Affairs*. A.B., University of Chicago; J.D., DePaul University College of Law.
- SUSIE AGNEW, *Assistant Director of Student Services*. Clark College.
- WILLIAM J. BAKER, *Assistant Director of Technology Services*.
- SUSAN BUSHUE-RUSSELL, *Assistant Director of Administration and Finance*. A.A.S., Lakeland College; B.S., Eastern Illinois University.
- KAREN H. MILLER, *Assistant Director for Admissions*. Midway College.
- LISA SCHRAGE, *Assistant Director for Donor Relations*. B.S., Marian College.

Faculty

- CYNTHIA M. ADAMS, *Clinical Professor of Law and Co-Director of the Latin American Law Program*. B.A., Kentucky Wesleyan College; J.D., Indiana University School of Law—Indianapolis.

- JUDITH FORD ANSPACH, *Professor of Law and Director, Ruth Lilly Law Library*. B.S., M.L.S., Kent State University; J.D., Mississippi College School of Law.
- CYNTHIA BAKER, *Clinical Associate Professor of Law and Director, Program on Law and State Government*. B.A., J.D., Valparaiso University.
- GERALD L. BEPKO, *Indiana University-Purdue University—Indianapolis Chancellor Emeritus, Indiana University Trustee Professor and Professor of Law*. B.S., Northern Illinois University; J.D., ITT/Chicago-Kent College of Law; LL.M., Yale Law School.
- SHAWN BOYNE, *Associate Professor of Law and Dean's Fellow*. B.A., Cornell University; M.B.A., University of Minnesota, J.D., University of Southern California's Gould School of Law, M.A., University of Wisconsin, L.L.M., Justus-Liebig-Universität.
- KAREN E. BRAVO, *Professor of Law; John S. Grimes Fellow and Dean's Fellow*. B.A., The University of the West Indies; J.D., Columbia University School of Law; LL.M., New York University School of Law.
- ROBERT BROOKINS, *Professor of Law*. B.S., University of South Florida; J.D., Ph.D., Cornell University.
- KENNETH D. CHESTEK, *Clinical Associate Professor of Law*. B.A., Pennsylvania State University; J.D., University of Pittsburgh School of Law.
- DANIEL H. COLE, *R. Bruce Townsend Professor of Law*. A.B., Occidental College; A.M., University of Chicago; J.D., Lewis and Clark Law School; J.S.M., J.S.D., Stanford Law School.
- JEFFREY O. COOPER, *Associate Professor of Law*. A.B., Harvard University; J.D., University of Pennsylvania Law School.
- ERIC DANNENMAIER, *Associate Professor of Law and Dean's Fellow*. B.A., Drury College; J.D., Boston University; LL.M., Columbia University; M. St., Oxford University.
- JAMES D. DIMITRI, *Clinical Professor of Law*. B.S., Indiana University; J.D., Valparaiso University School of Law.
- JENNIFER A. DROBAC, *Professor of Law and Director, Central and Eastern European Law Summer Program*. B.A., M.A., Stanford University; J.D., J.S.D., Stanford Law School.
- GEORGE E. EDWARDS, *Carl M. Gray Professor of Law; Faculty Advisor to the LL.M. Track in International Human Rights Law* (term ending, Spring 2011); *Director, Program in International Human Rights Law and John S. Grimes Fellow*. B.A., North Carolina State University; J.D., Harvard Law School.
- FRANK EMMERT, *John S. Grimes Professor of Law and Executive Director, Center for International and Comparative Law*. Erstes Juristisches Staatsexamen (J.D.), University of Munich Law School; LL.M., The University of Michigan Law School; Ph.D., University of Maastricht; Diploma, European University Institute.
- NICHOLAS GEORGAKOPOULOS, *Harold R. Woodard Professor of Law*. Ptyhion Nomikis, Athens University School of Law; LL.M., S.J.D., Harvard Law School.
- CARRIE HAGAN, *Clinical Associate Professor of Law*. B.A., University of Kansas; J.D., University of Cincinnati, College of Law.
- JOHN LAWRENCE HILL, *Professor of Law; Adjunct Professor of Philosophy and John S. Grimes Fellow*. B.A., Northern Illinois University; J.D., Ph.D., Georgetown University.
- LINDA KELLY HILL, *M. Dale Palmer Professor of Law*. B.A., J.D., University of Virginia.
- MAX HUFFMAN, *Associate Professor of Law and Dean's Fellow*. B.A., Cornell University, J.D., University of Chicago College of Law.
- LAWRENCE A. JEGEN, III, *Thomas F. Sheehan Professor of Tax Law and Policy*. A.B., Beloit College; J.D., M.B.A., The University of Michigan; LL.M., New York University School of Law.
- ROBERT A. KATZ, *Professor of Law and John S. Grimes Fellow*. A.B., Harvard College; J.D., University of Chicago Law School.
- ELEANOR D. KINNEY, *Hall Render Professor of Law, Co-Director of the William S. and Christine S. Hall Center for Law and Health*. B.A., Duke University; M.A., University of Chicago; J.D., Duke University School of Law; M.P.H., University of North Carolina.
- ANDREW R. KLEIN, *Paul E. Beam Professor of Law; Chief of Staff, Office of the Chancellor of Indiana University-Purdue University—Indianapolis*. B.A., University of Wisconsin; J.D., Emory University School of Law.
- NORMAN LEFSTEIN, *Professor of Law and Dean Emeritus*. LL.B., University of Illinois College of Law; LL.M., Georgetown University Law School.
- MARÍA PABÓN LÓPEZ, *Professor of Law and Dean's Fellow*. B.A., Princeton University; J.D., University of Pennsylvania Law School.

GERARD N. MAGLIOCCA, *Professor of Law; Co-Director, Chinese Law Summer Program and John S. Grimes Fellow*. B.A., Stanford University; J.D., Yale Law School.

ALLISON MARTIN, *Clinical Associate Professor of Law*. B.S., J.D., University of Illinois.

DEBORAH MCGREGOR, *Clinical Professor of Law and Assistant Director of Legal Analysis, Research and Communication*. B.A., University of Evansville; J.D., Georgetown University Law School.

SUSANAH M. MEAD, *Professor of Law*. B.A., Smith College; J.D., Indiana University School of Law—Indianapolis.

EMILY MORRIS, *Associate Professor of Law and Dean's Fellow*. A.B., Harvard University; J.D., University of Michigan Law School.

NOVELLA NEDEFF, *Clinical Associate Professor of Law*. B.A., J.D., Indiana University.

DAVID ORENTLICHER, *Samuel R. Rosen Professor of Law and Co-Director of the William S. and Christine S. Hall Center for Law and Health*. A.B., Brandeis University; J.D., M.D., Harvard College.

JOANNE ORR, *Clinical Professor of Law*. B.S., Indiana State University; J.D., California Western School of Law.

ANTONY PAGE, *Associate Professor of Law and Dean's Fellow*. B. Comm., McGill University; M.B.A., Simon Fraser University; J.D., Stanford Law School.

MICHAEL J. PITTS, *Associate Professor of Law; Dean's Fellow and John S. Grimes Fellow*. B.S.J., Northwestern University; J.D., Georgetown University Law Center.

PETER A. PRESCOTT, *Associate Professor of Law*. B.A., Augustana College; M.S., The University of Chicago; M.P.A., The University of Texas at Austin; J.D., The University of Houston Law Center.

FRAN QUIGLEY, *Visiting Professor of Law*. B.A., Hanover College; M.A., Indiana University; J.D., Indiana University School of Law—Indianapolis.

FLORENCE WAGMAN ROISMAN, *William F. Harvey Professor of Law*. B.A., University of Connecticut; LL.B., Harvard Law School.

JOAN M. RUHTENBERG, *Professor of Law and Director of Legal Analysis, Research and Communication*. B.A., Mississippi University for Women; J.D., Indiana University School of Law—Indianapolis.

JOEL M. SCHUMM, *Clinical Professor of Law and Director, Judicial Externship Program*. B.A., Ohio Wesleyan University; M.A., University of Cincinnati; J.D., Indiana University School of Law—Indianapolis.

CARLTON M. WATERHOUSE, *Associate Professor of Law and Dean's Fellow*. B.S., Pennsylvania State University; J.D., Howard University School of Law; M.T.S., Emory University, Chandler School of Theology; Ph.D., Emory University.

FRANCES WATSON, *Clinical Professor of Law*. B.S., Ball State University; J.D., Indiana University School of Law—Indianapolis.

LLOYD T. WILSON, JR., *Professor of Law and Co-Director, Chinese Law Summer Program*. B.A., Wabash College; M.A., Duke University; J.D., Indiana University School of Law—Bloomington.

MARY T. WOLF, *Clinical Professor of Law and Director of Clinical Programs and Externships*. B.A., Saint Xavier College; J.D., University of Iowa College of Law.

R. GEORGE WRIGHT, *Lawrence A. Jegen III Professor of Law*. A.B., University of Virginia; Ph.D., Indiana University; J.D., Indiana University School of Law—Indianapolis.

Emeriti Faculty

THOMAS B. ALLINGTON, *Professor of Law Emeritus*. B.S., J.D., University of Nebraska; LL.M., New York University School of Law.

EDWARD P. ARCHER, *Professor of Law Emeritus*. B.M.E., Renesselaer Polytechnic Institute; J.D., LL.M., Georgetown University Law School.

JAMES F. BAILEY, III, *Professor of Law Emeritus and Director of the Law Library Emeritus*. A.B., J.D., M.A.L.S., The University of Michigan.

CLYDE HARRISON CROCKETT, *Professor of Law Emeritus*. A.B., J.D., University of Texas; LL.M., University of London (The London School of Economics and Political Science).

DEBRA A. FALENDER, *Professor of Law Emerita*. A.B., Mount Holyoke College; J.D., Indiana University School of Law—Indianapolis.

DAVID A. FUNK, *Professor of Law Emeritus*. A.B., College of Wooster; J.D., Case Western Reserve University School of Law; M.A., The Ohio State University; LL.M., Case Western Reserve University; LL.M., Columbia Law School.

PAUL J. GALANTI, *Professor of Law Emeritus*. A.B., Bowdoin College; J.D., University of Chicago Law School.

HELEN P. GARFIELD, *Professor of Law Emerita*. B.S.J., Northwestern University; J.D., University of Colorado School of Law.

HAROLD GREENBERG, *Professor of Law Emeritus*. A.B., Temple University; J.D., University of Pennsylvania Law School.

JEFFREY W. GROVE, *Professor of Law Emeritus*. A.B., Juniata College; J.D., George Washington University Law School.

WILLIAM F. HARVEY, *Carl M. Gray Professor of Law & Advocacy Emeritus*. A.B., University of Missouri; J.D., LL.M., Georgetown University Law School.

W. WILLIAM HODES, *Professor of Law Emeritus*, A.B., Harvard College; J.D., Rutgers University School of Law—Newark.

WILLIAM ANDREW KERR, *Professor of Law Emeritus*. A.B., West University; B.S., Duke University; J.D., LL.M., Harvard Law School.

WILLIAM E. MARSH, *Professor of Law Emeritus*. B.S., J.D., University of Nebraska.

RONALD W. POLSTON, *Professor of Law Emeritus*. B.S., Eastern Illinois University; LL.B., University of Illinois College of Law.

KENNETH M. STROUD, *Professor of Law Emeritus*. A.B., J.D., Indiana University—Bloomington.

JAMES W. TORKE, *Carl M. Gray Professor of Law Emeritus*. B.S., J.D., University of Wisconsin.

JAMES PATRICK WHITE, *Professor of Law Emeritus*. A.B., University of Iowa; J.D., LL.M., George Washington University Law School.

LAWRENCE P. WILKINS, *William R. Neale Professor of Law Emeritus*. B.A., The Ohio State University; J.D., Capitol University Law School; LL.M., University of Texas School of Law.

Law Library Faculty

JUDITH FORD ANSPACH, *Professor of Law and Director, Ruth Lilly Law Library*. B.S., M.L.S., Kent State University; J.D., Mississippi College School of Law.

DEBRA DENSLAW, *Reference Librarian*. B.A., Franklin College; M.S., University of Illinois, Urbana-Champaign; J.D., Valparaiso School of Law.

RICHARD HUMPHREY, *Reference Librarian*. A.A., Brewton-Parker Junior College; B.A., Georgia Southwestern College; M.L.S., University of Kentucky.

WENDELL E. JOHNTING, *Assistant Director for Technical Services*. A.B., Taylor University; M.L.S., Indiana University.

CATHERINE LEMMER, *Head of Information Services*. B.A., Lawrence University; J.D., University of Wisconsin; M.S., University of Illinois.

CHRIS E. LONG, *Cataloging Librarian*. B.A., Indiana University; M.A., Indiana University; M.L.S., Indiana University.

STEVEN R. MILLER, *Reference Librarian*. B.S., Lebanon Valley College of Pennsylvania; M.A., Dominican University; M.S., Northwestern University; J.D., The John Marshall Law School.

MAHNAZ K. MOSHFEGH, *Serials/Acquisitions Librarian*. B.A., National University of Iran; M.S., Tehran University; M.A., Ball State University; M.L.S., Ph.D., Indiana University.

MIRIAM A. MURPHY, *Associate Director*. B.A., Purdue University; J.D., M.L.S., Indiana University—Bloomington.

KIYOSHI OTSU, *Automated Services and Media Librarian*. A.A., Parkland College; A.B., M.S., C.A.S., University of Illinois.

Indiana Law Review

Volume 44

2011

Number 2

Copyright © 2011 by the Trustees of Indiana University

TABLE OF CONTENTS

TRIBUTES

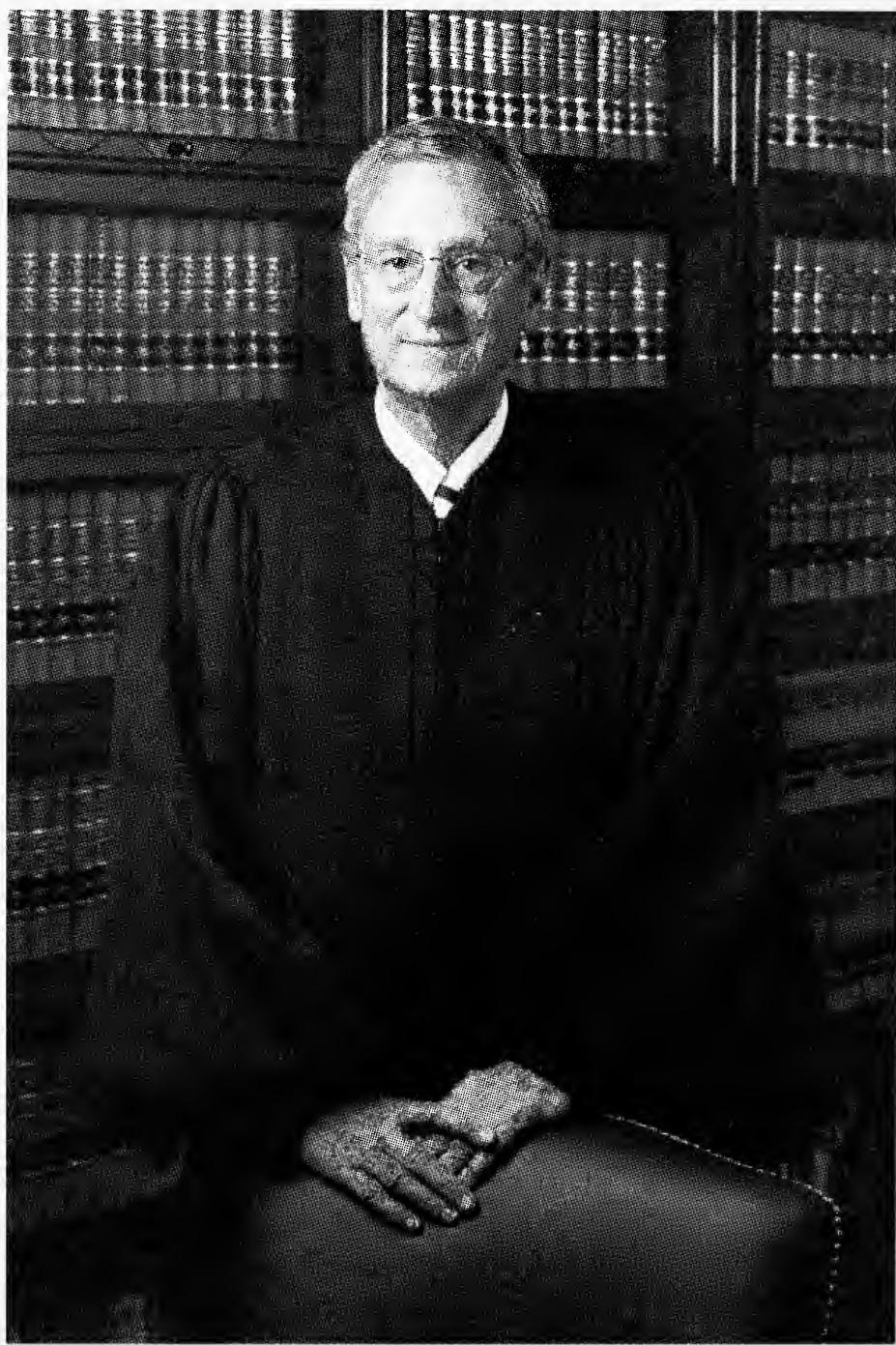
A Tribute to Justice Theodore Boehm	<i>Gerald L. Bepko</i>	341
Tribute to Justice Theodore R. Boehm	<i>Joel M. Schumm</i>	347
	<i>Paul L. Jefferson</i>	
Professor Henry C. Karlson		
A Remembrance		
October 29, 2010	<i>William F. Harvey</i>	353

ARTICLES

Judicial Strategy and Legal Reason	<i>Evelyn Keyes</i>	357
A Case of Doubtful Certainty: The Court Relapses into Search Incident to Arrest Confusion in <i>Arizona v. Gant</i>	<i>George M. Dery III</i>	395
Towards Recognizing and Reconciling the Multiplicity of Values and Interests in Trademark Law	<i>Michael S. Mireles, Jr.</i>	427
Wrongfully Incarcerated, Randomly Compensated—How to Fund Wrongful-Conviction Compensation Statutes	<i>Deborah Mostaghel</i>	503

NOTES

Who Needs Contract Law?—A Critical Look at Contractual Indemnification (or Lack Thereof) in FHAA and ADA “Design and Construct” Cases	<i>Charles Daugherty</i>	545
Application or Registration?: Confusion Regarding the Copyright Act’s Prerequisite to Copyright Infringement Lawsuits	<i>Thomas M. Landrigan</i>	581
Justice or Mental Health . . . Should Litigants Have to Choose? Mental Health as a Reason to Proceed Anonymously	<i>Sarah Orme</i>	605
“Growing Pains” in Indiana Age Discrimination Law	<i>David T. Vlink</i>	627



JUSTICE THEODORE R. BOEHM

Indiana Law Review

Volume 44

2011

Number 2

TRIBUTES

A TRIBUTE TO JUSTICE THEODORE BOEHM

GERALD L. BEPKO*

On the day I began to write this tribute to retired Indiana Supreme Court Justice Theodore Boehm, I received an email news item from the *Indianapolis Business Journal* reporting on a celebrated probate case. The case involved disputes over the estate of one of Indiana's most successful and wealthy business leaders. There had been some intra-family disagreements, and the family member who had been designated as the executor for the estate, after engaging in some questionable activities, was removed by the presiding judge. Because the assets of the estate were valued at more than two billion dollars, and because the issues raised were sensitive and complex, the judge had to appoint someone of special stature, competency, and judgment to serve as a new executor. The news release explained that the judge had appointed Ted Boehm to be the executor and that Ted had accepted the appointment.

This appointment made sense to me not only because Ted Boehm is so talented, experienced, and wise, but because the appointment is another episode in a life of extraordinary consequence—a life filled with a mixture of professional achievements, leadership in a broad range of roles in his relevant communities, a repeated willingness to answer the call to service in his home state of Indiana, and a confidence-inspiring demeanor and wit which are hallmarks of Ted's exceptional life. There are few, if any, who have contributed as much to the life of Indiana with such constancy and a special understanding of the culture and needs of the Hoosier state.

The foundation for his life in Indiana was a superior education. Like so many leaders of this era in Indiana's history, Ted Boehm studied at Shortridge High School before going off to one of the nation's great universities. In Ted's case, this was Brown, from where he graduated summa cum laude on his way to Harvard Law School. From Harvard he went to the U.S. Supreme Court, where he had the opportunity of clerking for Chief Justice Earl Warren. Prior to his appointment by President Dwight Eisenhower, Chief Justice Warren had been a presidential candidate and popular governor of California, so popular that in one of his reelection campaigns he was nominated by all of California's major political parties. He was not only the Chief Justice who had brought landmark

* Former Interim President of Indiana University; IUPUI Chancellor Emeritus; Indiana University Trustees Professor and Professor of Law; former Dean of the Indiana University School of Law—Indianapolis.

cases like *Brown v. Board of Education*¹ to a successful conclusion with a unanimous vote of the Justices, but at the time of Ted's clerkship, the Chief Justice was tapped by President Lyndon Johnson to head up the commission that was charged with studying and reporting on the assassination of President John F. Kennedy.

There are few people of that era who would not know instantly what was described by the words "Warren Commission." Its purposes were to gather the facts of the Kennedy assassination, to put this horrible event in perspective, and to help the nation move on. It was appointed on November 29, 1963, only seven days after President Kennedy's death, and filed its report on September 22, 1964. The time in between, during which the Commission conducted its investigation and prepared its report, was roughly the time when Ted clerked for the Chief Justice.

One could make the case that there was no more important year in the history of the office of the Chief Justice than 1963-64. Earning summa cum laude honors at Brown, a law degree from Harvard, and serving as a clerk in Earl Warren's chambers during a year when the eyes of the world were on the Warren Commission provided Ted with a remarkable background for nearly anything he would choose to do. If there was any better way in this era for a young lawyer to get an accelerated start toward a life of consequence, it is not easy to envision.

With a world of opportunities waiting, Ted Boehm answered the call of his adopted home state and returned to Indiana to join Baker and Daniels, one of Indiana's largest and most prominent law firms. At that time Ted probably could have chosen any threshold opportunity in any place of his liking, but he was influenced by John F. Kennedy's Inaugural Address from 1960. That was the speech in which the young President spoke of a new generation of leadership who should "ask not what your country can do for you; ask what you can do for your country."² Ted Boehm has lived this admonition, starting with his decision to return to Indiana and his choice of a life in a place where he would seek to make an important difference.

In this sense Ted seems to have contributed to a pattern for Baker and Daniels lawyers I know, such as Jack Swarbrick, who was a key leader of the sports movement in Indianapolis before going back to his alma mater, Notre Dame, to be athletic director. Ted was instrumental in recruiting Jack Swarbrick by serving as an example of the sort of person who had come back to Indianapolis and by making the case that Indianapolis was a city of the future with special opportunities. Jack said that none of his Stanford Law School classmates went to smaller cities like Indianapolis. They followed the allure of the largest firms located in a few of the largest population centers. Ted's example confirmed what Jack had heard from a priest friend about what was important in building a professional life—a priest friend who was on his way from an affluent parish in San Francisco to an inner city parish in Detroit. This influenced Jack to think about what was most important to a life of fulfillment and helped bring Jack to

1. 347 U.S. 483 (1954).

2. President John F. Kennedy, Inaugural Address (Jan. 20, 1961).

Indianapolis and to a kindred spirit who served as a mentor.

Another Baker and Daniels lawyer whose life was influenced by Ted is David Johnson, who now leads Indiana's efforts to expand economic opportunity and create jobs in the life sciences. Ted became a mentor to David while David was in college and Ted was the president of the Meridian-Kessler Neighborhood Association. David's father was pastor at First-Meridian Heights Presbyterian Church. In the 1960s and early 1970s, David's father and his clergy colleagues worked to establish the Meridian-Kessler Association to avoid racial segregation and block racially divisive lending practices in the Meridian-Kessler area of the city. Up until that time, Meridian-Kessler had been a stable and diverse family neighborhood. David saw Ted Boehm, who at the time was a young partner at Baker and Daniels, play a leadership role in this successful movement.

Later, when David clerked during law school at Baker and Daniels, Ted was his assigned mentor. Still later, when David joined the firm as an associate in 1983, fresh from his term as a Rhodes Scholar at Oxford and as a student at the Harvard Law School, Ted was the managing partner, the first person to serve in this leadership role. In subsequent years Ted and David became active together in Democratic Party politics, especially the campaigns of Frank O'Bannon for lieutenant governor and governor. David served as counsel and close advisor to Governor O'Bannon, another person with a fully developed sense of community and civic responsibility.

Ted's role as a champion of civic responsibility and values not only had an impact on those who considered him a mentor, but were also a factor in the way he practiced law as a highly ethical and capable lawyer who served his clients well. This is why he was chosen to argue two different cases before the U.S. Supreme Court. It also contributed to his emergence as a leader of the firm as its first managing partner, but there was more.

In a firm comprising scores of very talented and competitive peers, the person chosen as a leader must be as good as the best in terms of professional work and also be the type of person who has very good judgment and inspires trust and confidence among peers. One of the most easily observable aspects of Ted's life has been the enormous trust and confidence he has inspired in and among various constituencies. He lives by the highest standard of ethical conduct and is a living example of the legal concept known as good faith.

His years as the first Baker and Daniels managing partner were of considerable consequence. In collaboration with another leader of the sports movement and real estate developer, Michael Browning, Ted played a lead role in planning the new office space that the firm acquired in a new building at New York and Meridian Streets. Consistent with his modesty and egalitarian spirit, he arranged for offices in the new space that were all roughly the same size, in contrast to firms which have chosen a hierarchical pyramid of larger and larger offices for more senior partners. And he planned the space so that every office would have maximum external views. It was called a "train trestle"—all windows with no cubicles—built around the library and work areas in the center of the building.

During this same time, Ted played a crucial role in creating and advancing the sports movement in Indianapolis. He was active in and served as the first

chair of the board of the Indiana Sports Corporation (ISC), which has provided energy and leadership for Indianapolis's ambitions to make sports a driver of economic and community development. In these early years, the ISC managed Indianapolis's first large multi-sport event—the National Sports Festival. This event built momentum, gave confidence to the people of Indiana that the sports initiative would bring success, and began a pattern in other cities around the country of copying the Indianapolis experience.

Not long after the National Sports Festival, I had my first opportunity to work with Ted. It was in connection with the Pan American Games, which were held in Indianapolis in 1987. Ted was the chair and CEO of the Pan American Games Host Committee. Since many of the sports venues were on the IUPUI campus, where I worked, I had lots of dealings with Ted. From this vantage point it seemed clear to me that his leadership and passionate efforts were essential to the success of the Games.

Although there was a reservoir of volunteers left from the National Sports Festival in 1982, the Pan American Games volunteer effort in 1987 was of an entirely different and greater magnitude. It was a consummate example of a community, a city, coming together in a volunteer spirit to achieve something very important. It was an uplifting moment of unity. In all of this, Ted played the primary role—including leading on the largest issues, on which the success of the Games depended, and often on the smallest details that were essential to this successful community volunteer effort.

As IUPUI campus Chancellor, I was very impressed with Ted's work, and, inspired by his leadership, I made special efforts to support the Games. It was also a pleasure for me to serve for nearly twenty years on the board of the Indiana Sports Corporation and see the product of Ted's early leadership and the enormous benefit that it brought to Indianapolis and Indiana. Former UN World Food Program leader and current Indiana Pacers CEO Jim Morris, one of Indiana's most prominent leaders, has summed up Ted's contributions in this and other areas of Indiana's growth and development by saying that Ted "has an unusual commitment to simply want to make things better. When somebody has the passion and commitment that he has and also has the skills and talent and brainpower that he has, it's an incredible combination."

As Jim Morris's quote suggests, Ted's community leadership has not been confined to sports. He was and continues to be active in leading the development of the arts and cultural dimensions of our Indianapolis community. He was an early president of the Penrod Society, which hosts the Penrod Arts Fair each year, and he was one of the founding members of the Economics Club of Indianapolis, which presents monthly speakers of national significance on a range of public policy issues. He has been a member of the board of WFYI, our region's public broadcasting outlet, and his interest in downtown Indianapolis has been reflected in his board memberships for the National Art Museum of Sport and his very important service on the board of the Indianapolis Conventions and Visitors Association (ICVA). I served as the ICVA board chair for a few years, which makes me even more confident in the conclusion that this represents another of Ted's excellent contributions to the public good. Most recently, I've had the pleasure to serve on the Indianapolis Cultural Development Commission, which

Ted chairs. This entity supports the arts, cultural activities, and cultural tourism in Indianapolis, and this has given me more firsthand knowledge of the crucial importance of Ted's leadership.

Ted has achieved enough in his role as community leader in Indiana so that one might think he lived in Indianapolis continuously throughout his career as a lawyer. In fact, Ted left Indianapolis to work for several years as general counsel for two of General Electric's important subsidiaries, first in Louisville as counsel for the appliance division, and finally as counsel for the aircraft engine division in Cincinnati. After those years away, he returned to his home city to work briefly for his home firm, Baker and Daniels, and then to accept an appointment by Governor Evan Bayh to join the Indiana Supreme Court. He took his seat on the court in 1996 and retired in 2010 after having contributed in very important ways to the continuing development of the supreme court and the lower court system in Indiana.

As a law professor and dean, I have always been proud of the growing stature of the courts in Indiana. The quality of the Indiana judiciary in general, which has grown under the leadership of the supreme court by Chief Justice Randall Shepard, is a remarkable phenomenon. It's something all Hoosiers should understand and admire. By the time Ted joined the Indiana Supreme Court, it had already become well known for its high quality, thus making it even more difficult to add to its growing reputation. Ted managed to do that, however, with the basic quality of his intellectual contribution to the court's body of decisions and by bringing his management experience in corporate positions to bear on the work of the court. This led to such things as improvements in the way those eligible for jury service are identified and gathered into pools of candidates for selection.

For four years in retirement, I taught a course titled "Leadership and Law." The purpose of the course was to give law students an opportunity to think about leadership and learn the language and ideas that have been developed by those who study the subject. It was meant to give students a framework for thinking about leadership in all the ways that lawyers must lead: in their communities, in their firms, in their work to reform law in the public interest, in the protection of those who cannot afford full participation in the system of justice, in their efforts to continue to improve the profession, and in their nurturing of each new generation of lawyers. When the class talked about leadership in this context, there was often a reference to "impact" and how those who studied law could improve conditions in the world. This translated into a question: How do we define a life well lived grounded in law?

All the lawyer/leaders who came to speak to my class incorporated in some way these questions in their reflections on the subject matter. This was true of all those "grounded in law" who came to speak, whether it was U.S. District Judge Sarah Evans Barker, Center for Leadership Development CEO Dennis Bland, Indiana House of Representatives Speaker Brian Bosma, Clarian CEO Dan Evans, entrepreneur/author/publisher/lawyer/philanthropist Michael S. "Mickey" Maurer (for whom the Maurer School of Law at Indiana University—Bloomington is named), lawyer/bank CEO Robert McKinney, Indianapolis Mayor Bart Peterson, or Professors David Orentlicher and Florence Roisman.

Until I began to write this tribute, I had not realized that Ted Boehm was a mentor to two of those who spoke eloquently to my class on the subject of leadership: David Johnson and Jack Swarbrick. David and Jack were probably too young at the time to be affected by the Kennedy inaugural speech in the way that Ted Boehm was, but they were among those privileged to be mentored by Ted and imbued with the notion that serving good causes is life's highest calling. By way of their personal reflections, they certainly gave inspiration to the law students in my classes.

I'm sure Ted is proud of what they have done and the way they and others for whom Ted has served as mentor have chosen lives of meaning, adventure, and consequence. My reaction to learning of Ted's influence on the two of them is to conclude that if I teach this course again, I will not only continue to invite David Johnson and Jack Swarbrick to speak, but I will ask Ted if he would be willing to open the class as the first speaker. He may well be the most accomplished of all at defining a life well lived grounded in law and helping others like David and Jack to understand these elements of life and leadership. I doubt that there could be any better way to start the study of leadership and law or to give an orientation to young people entering the profession.

TRIBUTE TO JUSTICE THEODORE R. BOEHM

JOEL M. SCHUMM*
PAUL L. JEFFERSON**

We had the privilege of serving as law clerks to Justice Boehm and offer this tribute, with analogies to another of his passions,¹ to his service on the Indiana Supreme Court.

I. THE NUMBERS

Justice Boehm authored 485 majority opinions, 67 concurring opinions, and 81 dissenting opinions during his fourteen years on the Indiana Supreme Court. During most of those years, he wrote more opinions than any other justice.² Justice Boehm was also the go-to justice when an issue required a speedy resolution. For example, he authored, and the Court published, the opinion that allowed the lease of the Indiana Toll Road just one week after oral argument.³

II. THE BIG PLAYS

Justice Boehm's opinions often accomplish what only good lawyers can do: they make the complex seem simple. Each opinion begins with a short narrative meant to convey the central issue and holdings to lawyers and non-lawyers alike, and each opinion often takes several distinct and at times contradictory concepts and weaves them together in an analysis that, at the end, often appears self-evident. This was accomplished by Justice Boehm's intellect and drive for perfection, resulting in a work product many of his clerks referred to as his "patina."

Justice Boehm served on the Court when, after the Indiana Constitution was amended to change its jurisdiction,⁴ the Court was able to address a larger percentage of civil cases. Many of these issues, due largely to the Court's non-discretionary and mostly criminal docket, had not been addressed by the Court for some time. Justice Boehm was instrumental in updating and clarifying authority regarding various civil and constitutional issues. While this space is neither large enough to recite them all, nor is that the purpose of this tribute, a

* Clinical Professor of Law, Indiana University School of Law—Indianapolis. J.D., 1998, Indiana University School of Law—Indianapolis; M.A., 1994, University of Cincinnati; B.A., 1992, Ohio Wesleyan University.

** Partner, Barnes & Thornburg LLP. Indiana University School of Law—Indianapolis, J.D., 2002; Wabash College, A.B., 1992.

1. Justice Boehm was instrumental in helping make Indianapolis the amateur sports capital of the United States, not to mention he was a big fan of the Colts, Pacers, and Indians. *See generally* Joel M. Schumm, *Theodore R. Boehm*, in *JUSTICES OF THE INDIANA SUPREME COURT* 417, 417-18 (Linda C. Gugin & James E. St. Clair eds., 2011).

2. See IND. SUPREME COURT, 2009-2010 ANNUAL REPORT 7 (2010), available at <http://www.in.gov/judiciary/supremeadmin/docs/0910report.pdf>.

3. Bonney v. Ind. Fin. Auth., 849 N.E.2d 473 (Ind. 2006).

4. IND. CONST. art. VII, § 4.

few instances representing the depth and breadth of the issues bear mentioning.

In *In re WTHR-TV v. Cline*,⁵ Justice Boehm addressed various discovery and privilege issues and reminded practitioners that those who embark on discovery “fishing expeditions” should at least be aware of the “idea of the size, species, or edibility of the fish.”⁶ In *City of Gary ex rel. King v. Smith & Wesson Corp.*, Justice Boehm updated Indiana Supreme Court jurisprudence regarding nuisance law in a manner that has been deemed worthy to be a teaching tool for law students.⁷ In *R.L. McCoy, Inc. v. Jack*, the Court, via Justice Boehm, held that Indiana’s comparative fault regime does not allow a credit for severally liable settled defendants.⁸ *Cheatham v. Pohle* upheld Indiana’s statute requiring a percentage of a punitive damage award to be sent to the State.⁹ These examples are but a few.

Similarly, Justice Boehm’s tenure was marked by many landmark opinions regarding the Indiana Constitution. Analyzing special legislation will never be the same after *Municipal City of South Bend v. Kimsey*, which gave effect to article IV’s prohibition of certain special laws.¹⁰ The legislature’s ability to create and eliminate causes of action was confirmed in *McIntosh v. Melroe Co.*, which again upheld Indiana’s statute of repose in product liability cases.¹¹ And in *State v. Kimco of Evansville, Inc.*, the State’s takings analysis was updated.¹² Again, these are but a small sample.

In the criminal realm, Justice Boehm’s opinions addressed a wide range of subjects and offered invaluable guidance for the future. Within a couple of years of his appointment, he resolved the longstanding confusion regarding the proper time to litigate a claim of ineffective assistance of counsel, holding in *Woods v. State* that it could be brought only once and should usually be pursued in a post-conviction proceeding when a supporting record exists.¹³ More than a decade later, he wrote the opinion that rejected *Anders* briefs¹⁴ in Indiana, concluding that indigent defendants were “entitled to a review by the judiciary, not by overworked and underpaid public defenders.”¹⁵

Notable opinions involving search and seizure issues include *Litchfield v. State*,¹⁶ where the Court adopted more protective approach to Hoosiers’ trash,

5. 693 N.E.2d 1 (Ind. 1998).

6. *Id.* at 7.

7. See generally *City of Gary ex rel. King v. Smith & Wesson Corp.*, 801 N.E.2d 1222 (Ind. 2003).

8. *R.L. McCoy, Inc. v. Jack*, 772 N.E.2d 987, 989-91 (Ind. 2002).

9. *Cheatham v. Pohle*, 789 N.E.2d 467, 472 (Ind. 2003).

10. See generally *Mun. City of South Bend v. Kimsey*, 781 N.E.2d 683 (Ind. 2003).

11. *McIntosh v. Melroe Co.*, 729 N.E.2d 972, 980 (Ind. 2000).

12. See generally *State v. Kimco of Evansville, Inc.*, 902 N.E.2d 206 (Ind. 2009).

13. *Woods v. State*, 701 N.E.2d 1208, 1213-14 (Ind. 1998).

14. See generally *Anders v. California*, 386 U.S. 738, 744-45 (1967) (providing procedure to allow court-appointed counsel to withdraw from an appeal counsel believes is without merit).

15. *Mosley v. State*, 908 N.E.2d 599, 608 (Ind. 2009).

16. 824 N.E.2d 356 (Ind. 2005).

holding that

a search of trash recovered from the place where it is left for collection is permissible under the Indiana Constitution, but only if the investigating officials have an articulable basis justifying reasonable suspicion that the subjects of the search have engaged in violations of law that might reasonably lead to evidence in the trash.¹⁷

Justice Boehm also wrote for a unanimous Court in *Edwards v. State*,¹⁸ which held that “routine, warrantless strip searches of misdemeanor arrestees, even when incident to lawful arrests, are impermissible under the Indiana Constitution and the United States Constitution.”¹⁹

Finally, other decisions were grounded in the Court’s supervisory power to craft prospective rules for the more orderly operation of lower courts. In *Tyler v. State*,²⁰ the Court held that “a party may not introduce testimony via the Protected Person Statute if the same person testifies in open court as to the same matters.”²¹ In his final days on the Court, he authored a 3-2 opinion requiring trial courts to inform defendants who plead guilty that “an attorney is usually more experienced in plea negotiations and better able to identify and evaluate any potential defenses and evidentiary or procedural problems in the prosecution’s case.”²²

Justice Boehm also left a legacy not published in the *North Eastern Reporter*. He championed rule changes to make appellate practice easier for practitioners and judges and was instrumental in bringing the Indiana Supreme Court to the world by webcasting oral arguments. He also constantly considered and implemented ways to administer justice more efficiently, without sacrificing quality.

III. LEAGUE RELATIONS

Many of Justice Boehm’s opinions required the Court to address the actions of the other branches of state government. He always did so with respect. In *Kimsey*, decades of legislation based on population parameters that functioned as a legal sleight-of-hand were held no longer per se valid.²³ In *Bonney v. Indiana Financial Authority*, the controversial lease of Indiana’s toll road was upheld in a promptly issued, yet thorough, opinion.²⁴ In *Sholes v. Sholes*, the legislature’s attempt to force lawyers to work for free, no matter how laudable, was ruled

17. *Id.* at 357.

18. 759 N.E.2d 626 (Ind. 2001).

19. *Id.* at 627-28.

20. 903 N.E.2d 463 (Ind. 2009).

21. *Id.* at 465.

22. *Hopper v. State*, 934 N.E.2d 1086, 1088 (Ind. 2010).

23. See generally *Mun. City of South Bend v. Kimsey*, 781 N.E.2d 683 (Ind. 2003).

24. *Bonney v. Ind. Fin. Auth.*, 849 N.E.2d 473, 476 (Ind. 2006).

unconstitutional.²⁵

In sorting through the meaning of “percent” after the Indiana General Assembly amended the operating while intoxicated statute, Justice Boehm drew on humor and mathematics in explaining in *Sales v. State* that a contrary reading “would long since have produced not an impaired driver but a corpse, indeed one perhaps needing no embalming.”²⁶ In sum, Justice Boehm did not shirk the role a judge must perform, and he always acted with respect for the other branches. Indeed, Governor Mitch Daniels remarked upon Justice Boehm’s retirement that his “cooperative and practical temperament serves as a model for any who might come after him.”²⁷

IV. A TEAM PLAYER

Most opinions during Justice Boehm’s service were unanimous. When he disagreed, though, he also did so respectfully and often memorably. Perhaps no better example exists than his concurrence in *In re Wilkins*,²⁸ a rehearing opinion he reluctantly joined because if he did not, then the punishment would be worse. Relying on the wisdom of Lewis Carroll, he noted what would happen if he continued to dissent:

Lewis Carroll would love that result: half the Court believes no sanction is appropriate, and half would impose a small sanction, so the result is a major penalty. Only those who love the law could explain that to their children. To free parents everywhere from that burden, I concur in the result of granting rehearing . . .²⁹

His dissents in death penalty cases were especially significant,³⁰ as both

25. *Sholes v. Sholes*, 760 N.E.2d 156, 159 (Ind. 2001).

26. *Sales v. State*, 723 N.E.2d 416, 421 (Ind. 2000).

27. Mitch Daniels, Governor, Remarks at the Retirement Ceremony to Honor Indiana Supreme Court Justice Theodore R. Boehm (Sept. 30, 2010), available at http://media.ihets.org/vod/supremecourt/wmv/boehm_retirement.wmv.

28. 782 N.E.2d 985 (Ind. 2003).

29. *Id.* at 988 (Boehm, J., concurring).

30. *State v. Azania*, 865 N.E.2d 994, 1010 (Ind. 2007) (Boehm, J., dissenting) (concluding that “further pursuit of the death penalty [after twenty-five years] . . . violates the Indiana Constitution by imposing punishment that is both cruel and unusual”); *State v. McManus*, 868 N.E.2d 778, 792 (Ind. 2007) (Boehm, J., dissenting) (concluding that “the majority’s review of the evidence does not give sufficient deference to the trial court’s finding of mental retardation”); *Baird v. State*, 833 N.E.2d 28, 33 (Ind. 2005) (Boehm, J., dissenting) (“In short, I think it is plain that Baird is insane by any ordinary understanding of that term.”); *Williams v. State*, 793 N.E.2d 1019, 1031 (Ind. 2003) (Boehm, J., dissenting) (“This is a death penalty case, and Williams seems to me to have presented a plausible claim that DNA testing would present a reasonable possibility of affecting the decision of a jury to recommend the death penalty.”); *Daniels v. State*, 741 N.E.2d 1177, 1191 (Ind. 2001) (Boehm, J., dissenting) (“I believe the majority places too high a premium on finality and discounts evidence that suggests Daniels may not have been the perpetrator of these

Democratic and Republican governors later granted clemency to three death row inmates relying at least in part on his dissent or its rationale.³¹

V. A MENTOR TO NEW PLAYERS

Justice Boehm's remarkable work on the Court also left a mark on about twenty-five new law graduates who had the privilege of beginning their legal careers by learning from a brilliant, patient, and generous teacher. Justice Boehm taught clerks how to take complicated issues and explain them in a way that was accessible, engaging, and just. He did this not only by scribbling editing marks in the margins of multiple drafts, but also by frequently discussing the issues from when the case was first assigned to him until it was handed down. Clerks learned that every word had meaning, and the seemingly innocuous placement of a word could make an enormous difference.³² Clerks also learned the art of how to think about the law, and more specifically, how to combine a mixture of facts and legal doctrines in a way that allow for their distillation. Justice Boehm ran his chambers in a way that allowed new lawyers to begin their careers by learning the skill of how to think about the rule of law, and he inspired us to pursue careers that we hope allow us to leave the law better than we found it, as his tenure on the Court did.

horrendous crimes.”).

31. *Schumm, supra* note 1, at 419.

32. The most notable example is “only.” As clerks were told, the placement of the word can completely change the meaning of a sentence, as made clear in this example: “Only she loves me.” “She only loves me.” “She loves only me.”



HENRY C. KARLSON

PROFESSOR HENRY C. KARLSON*

A REMEMBRANCE

October 29, 2010

WILLIAM F. HARVEY**

An email from Mr. Thomas Doehrman, President of the Indiana Trial Lawyers Association, said, "I am very sorry to receive this news. Professor Carlson was a wonderful educator who was not afraid to speak his mind on a topic. He will be missed by the entire legal community." Mr. Doehrman was correct. Henry was a wonderful educator, and he will be missed by the entire legal community.

As the dean of the law school, I was pleased to invite Henry to join the faculty. I was very pleased to strongly support his request for tenure; it was a special moment because Professor Henry C. Carlson's law school professorship was masterful. He had an extraordinary understanding of his subjects. Additionally, he was a superb classroom professor and, therefore, could teach the information he knew. His exceptional personal energy sustained him. In his classes, each student discovered a level of learning that the student had not thought possible or could not have anticipated.

Henry understood the art of teaching law. Great law school teaching is much more demanding than conducting a "dialogue" in the setting of a casebook. That is merely an initial level, or a professorial entry level, from which a professor should ascend in not less than a year or perhaps two. The professor must master the body of information and know much more than the assigned casebook, a code, or a body of statutory law. That information must be organized into a classroom format. Then it is presented, if need be, using five or six different methods of instruction during one class period. A great professor knows and can use all of those methods as if they were a seamless web. Professor Carlson was a great professor. He was also devoted to Indiana University. He served on the Indianapolis Faculty Council Executive Committee and was the parliamentarian for the entire faculty council and for the school of nursing.

Professor Carlson was highly regarded in the legal profession as well. He taught in the venues of continuing legal education and the Indiana bar review course. He lectured to the Indianapolis Law Club for over twenty years. He was a member of the board of examiners of the National Board of Trial Advocacy. He was honored by the National Association of Social Workers and the U.S.

* Professor Henry C. Carlson taught for thirty-three years at Indiana University School of Law—Indianapolis. Courses he taught included criminal law, evidence, trial practice, trial advocacy, and seminars related to child abuse. He died on October 25, 2010 after a battle with cancer.

** Dean Emeritus and Carl M. Gray Professor of Law Emeritus at Indiana University School of Law—Indianapolis. This tribute essay is adapted from Dean Emeritus William F. Harvey's eulogy, which was delivered at Professor Carlson's funeral on October 29, 2010.

Food and Drug Administration. He had a keen interest in preventing child abuse and in the agencies that assist those situations. Thus, he distinguished the University of Illinois, from which he graduated, and Indiana University, to which he was dedicated.

His personality was radiant. The beginning of each day was joyful. After I left the deanship, we shared offices in the same suite. When he arrived in his office early in the morning, he remained for an hour or more. I would hear his door open and then slam closed. His walk in the corridor was similar to a march—it was fast, strong, clear, and fleet of foot. He was off to find a few students to challenge, to hear their views, and to offer his correction to something they might say. He loved the exchange. Moreover, his office door was always open to students. Whether a student was in his class or not, the student was welcome. If a student wanted information, he gave it. If a student wanted political debate, he gave that, too.

When I think about Henry, my mind turns to one piece of music in particular. With its trumpets and bouncing tempo, it is the third movement of Bach's Brandenburg Concerto Number Two in F Major. The piece summons images of a carefree nobleman who is dancing. Usually, it carries the instruction "allegro assai," Italian for very joyful and fast. I hope it is unnecessary to say that in using that musical metaphor, I am not describing how he danced. I am describing his mental and intellectual agility, his wit, and his patterns of thought, some of which came so fast that they were similar to a fast and joyful dance.

Professor Karlson was different from other faculty members. If he was in the law school building for a ten-hour day (or longer), he taught for ten hours. After his formal classes, if students were not in his office, he went to them. Wherever he was in the building, he taught constantly. During the whole of each day, he never stopped. I thought of him as "our ten-hour man." If he could not find a student, then he might stop and lecture another member of the faculty. As the dean, I did not object or intervene. If Henry lectured a student or another faculty member, at the end of the day each knew more than he knew when the day began—although a faculty member might be very reluctant to admit it.

One of his courses was evidence. Evidence is one of the most important courses a law school can offer because its principles are some of the common law's greatest concepts. They develop the transcending standards of fair trials, reasonable decisions made without bias, mandates issued upon a civilized and acceptable basis, the right to give evidence to a neutral forum, and the right to confront adverse claims or evidence. It is an extraordinary person who understands this.

Professor Henry C. Karlson was an extraordinary person. The beacons in his life were sustaining principles, those principles that maintain the social order and allow one to be impervious to defeat. They give much more. If they are known and used, they allow you to live a life of prominence and consequence. They allow you to sustain a peaceful and constructive social order even after your own life ends.

Thomas Babington Macaulay expressed very similar thoughts in this verse:

To every man upon this earth
Death cometh soon or late.
And how can man die better
Than facing fearful odds,
For the ashes of his fathers,
And the temples of his Gods . . .¹

Macaulay's verse was an evocation of the Roman Empire's grandeur. But it was much more. It speaks about transcending principles that guide man in creating a great social order and principles that guide man in defending a great social order.

For another reason, I mention Macaulay's verse today: Henry was a superb Roman scholar. He well understood the Roman Senate, the writings of Cicero and his defense of the Roman Republic, and the wars with Carthage. If you asked him a question about Rome, his first answer was a question—"Do you mean Rome in the days of the Republic, or in the days of the Empire?"—and then he offered an analysis of each.

Macaulay's verse, as Henry knew, meant that great principles, similar to a constellation of stars, sustain a collection of values. Pursuant to those values, you can align your life and also know that they remain after life's end. As a result, they are your most important estate, your most important gift to your family and to the generations that will trace their existence to you. They are greater than items of property—the corporeal remnants of a life—because they give peace, harmony, social tranquility, creativity, respect, and reverence.

Henry vigorously defended those principles. He took wholesome joy from their American fulfillment and from seeing the system whole, springing season after season from the trampoline of values and moral order which are the very warp and woof of freedom and progress. Furthermore, in his entire advocacy, I never heard him offer one unkind word about another person. Did he disagree with others? Yes, constantly, but it was always a principled disagreement, and it was never in vilification or defamation.

Henry lived Macaulay's verse because he vigorously protected our principles when facing fearful odds. He volunteered to fight in Vietnam—yes, he volunteered. He was a member of the famous 101st Airborne Division. In Vietnam, he was awarded the Bronze Star. He became a judge in trial by court-martial. Later, he became the chief of military justice for the 101st Airborne Division. He returned to the United States and was assigned to write the after-action report of the highly publicized My Lai trials.

When we spoke about that war, we understood each other. When I told him that my cousin, a Marine officer who graduated from Indiana University and was from Fort Wayne, Indiana, was killed in action in Vietnam, he understood. That requires a special understanding, and Henry had that special understanding. When I said that in the earlier 1950s in the U.S. Navy, I had firsthand experience

1. THOMAS BABINGTON MACAULAY, ESSAYS, CRITICAL AND MISCELLANEOUS 543 (1846).

in Indochina when the French were there, he understood. Again, that requires a special understanding—and again, Henry had that special understanding.

Over one hundred years after Macaulay's verse, and almost sixty years before today, Justice Felix Frankfurter expressed the proposition that the law, more than any other calling, has been concerned with those standards, criteria, and appeals to right and reason that have had a dominant share in begetting a civilized society. That was Professor Henry C. Karlson's understanding. It was his commitment. It was his instruction to students. He taught so that they, too, could share in begetting a civilized society. This is why he stayed the course, why he fought the good fight, and why he kept the faith.

Henry's life was joyful and fast. It was also very short. Henry, Nancy, and their family richly deserved another twenty years together. After them, all of us did.

This afternoon, Henry will be placed in a special area of the Crown Hill Cemetery: the "Field of Valor." It is the veterans' section, where the flag always flies, an eternal flame burns, and gratitude is always expressed. I shall always remember that he is there.

I shall always remember Henry in another way. Now he is among the stars. There in the darkened sky, I find his star. It is so bright that I shall love the night and rue the garish sun.

Go now, my gallant, gentle friend, my colleague, my brother in arms—go and be with God.

Indiana Law Review

Volume 44

2011

Number 2

ARTICLES

JUDICIAL STRATEGY AND LEGAL REASON

EVELYN KEYES*

ABSTRACT

Contemporary legal philosophers generally reject traditional legal reasoning as mere “conventionalism.” They argue that resort must be had to judicial strategies derived from moral and political theory—such as originalism, moral constructivism or perfectionism, minimalism, or pragmatism—to solve “hard cases,” particularly in divisive and developing areas of constitutional law. In this paper, I argue that the rejection of traditional legal reasoning is misdirected and destructive. Legal reason, as commonly understood and employed by traditionalist judges, is neither mere conventionalism nor one among many types of judicial strategies—and it is a particularly inadequate one where “hard cases” are concerned. Rather, it is a dynamic form of practical reason or applied moral reason operating within a flexible system of ordered liberties that is the positive law. It can resolve all cases, no matter how complex or novel, and it functions to maintain the purpose, integrity, and functionality of the rule of law over time. All judges should decide cases traditionally—not by the constructive use of non-traditional judicial strategies.

INTRODUCTION

Philosophers and judges generally agree on what legal reason is. Everyone knows that at least in most cases, judges reason deductively from conventionally fixed principles in the positive law¹ as applied to the facts of particular cases to reach judgments that prescribe the rights and obligations of the parties in the case and in similar future cases—and that is traditional “legal reasoning.” In the terminology of Ronald Dworkin, it is “conventionalism,”² and it is closely linked

* Justice, Texas Court of Appeals, First District. M.A., Ph.D., Rice University; M.A., Ph.D., University of Texas; J.D., University of Houston Law Center; B.A., Tulane University.

1. I define the “positive law” as the complex of officially generated, approved, and enforced constitutional principles, statutes, rules, and precedents upon which American judges generally rely to decide cases.

2. See RONALD DWORAKIN, LAW’S EMPIRE 114-17 (1986) [hereinafter DWORAKIN, LAW’S EMPIRE] (describing traditional jurisprudence as mere “conventionalism” and arguing that it cannot provide any justification for the resolution of issues that have not been settled one way or the other

to legal positivism, or the theory “that a community’s law consists only of what its lawmaking officials have declared to be the law.”³ This common knowledge then serves as the springboard for what I will call the “anti-traditionalist” argument. The anti-traditionalist argument states that traditional or “conventional” legal reason operates as common knowledge says it does, and that it works well enough in non-controversial, settled areas of the law, but because it cannot resolve “hard cases” for which the positive law provides no clear answer, judges must go outside the positive law to fill in the gaps.⁴ This argument is accepted by legal philosophers as diverse as Dworkin⁵ and Richard Posner,⁶ and it opens the door to alternative judicial strategies for solving “hard cases” that proponents variously claim reflect logical deductions from the objectively true moral propositions that at the highest level guide sound legal

by whatever institutions have conventional authority to decide them).

3. RONALD DWORKIN, JUSTICE IN ROBES 187 (2006) [hereinafter DWORKIN, JUSTICE]. Dworkin defines legal positivism as the thesis “that a community’s law consists only of the explicit commands of legislative bodies.” *Id.* at 212.

4. DWORKIN, LAW’S EMPIRE, *supra* note 2, at 130-50; cf. Alex Kozinski, *What I Ate for Breakfast and Other Mysteries of Judicial Decision Making*, in JUDGES ON JUDGING: VIEWS FROM THE BENCH 76 (David M. O’Brien ed., 2004) (describing traditional legal reasoning). Judge Kozinski explains:

[T]here are more or less objective principles by which the law operates, principles that dictate the reasoning and often the result in most cases. . . . Now, these principles are not followed by every judge in every case, and even when followed, there is frequently some room for the exercise of personal judgment.

But none of this means principles don’t exist or that judges can use them interchangeably or ignore them altogether. Let me give you an example of one principle I think is extremely important: *Language has meaning*. This doesn’t mean every word is as precisely defined as every other word, or that words always have a single, immutable meaning. What it does mean is that language used in statutes, regulations, contracts and the Constitution place an objective constraint on our conduct. The precise line may be debatable at times, but at the very least the language used sets an outer boundary that those interpreting and applying the law must respect. When the language is narrowly drawn, the constraints are fairly strict; when it is drawn loosely they’re more generous, but in either case they do exist. . . .

Another very important principle is that judges must deal squarely with precedent. . . . Precedent, like language, frequently leaves room for judgment. But there is a difference between judgment and dishonesty, between distinguishing precedent and burying it.

Id. at 78-79.

5. See generally DWORKIN, LAW’S EMPIRE, *supra* note 2; Ronald Dworkin, *Darwin’s New Bulldog*, 111 HARV. L. REV. 1718 (1998).

6. See RICHARD A. POSNER, THE PROBLEMATICS OF MORAL AND LEGAL THEORY vii-viii, 240-52 (1999) [hereinafter POSNER, PROBLEMATICS]; Richard A. Posner, *The Problematics of Moral and Legal Theory*, 111 HARV. L. REV. 1637, 1694 (1998).

judgment,⁷ lead to the socially “best” results for the community,⁸ or reflect the reconstructed “original intent” of the drafters of the law.⁹

The anti-traditionalist argument has been so successful that Dworkin indicates that “[t]he political influence of legal positivism has sharply declined in the last several decades . . . and it is no longer an important force either in legal practice or in legal education.”¹⁰ “Government,” he assures us, “has become too complex to suit positivism’s austerity.”¹¹ Legislative codes can no longer “purport to supply all the law that a community needs” when “technological change and commercial innovation outdistance the supply of positive law,” and thus, the idea has “steadily gained in popularity and in constitutional practice that the moral rights people have against lawmaking institutions have legal force.”¹² And, he might also have argued, the alternative strategies to traditional legal reasoning that compete with his moral reading of the Constitution have similarly gained force in legal theory and practice, pushing out traditional jurisprudence as insufficient to serve as a practice guide for sound judicial problem solving and unworthy even of scholarly attention.

But what if the anti-traditionalist portrait of traditional jurisprudence is wrong? What if legal reason as employed by traditionalist judges is not correctly described as mere conventionalism or legal positivism? Suppose that it is better understood as a form of applied moral reason, or practical reason, operating within the constraints of the positive law that is integral to maintaining the fairness, purpose, integrity, and functionality of the rule of law. But if we suppose that we have been missing the boat on understanding legal reason and then turn to actually parsing out how traditional jurisprudence works and what the process of legal reason upon which it relies actually is, what happens to the justification for rejecting traditional legal reasoning in favor of alternative strategies? And, worse, what if adherence to sound legal reasoning is essential to the integrity and functionality of the law, and the alternative strategies now being taught and recommended for the practice of our courts are actually blueprints for judicial advocacy that undermine the systemic integrity of the law that traditional jurisprudence protects? What happens to our legal system when judges become advocates for social justice, for the “best” construction of moral principle, or for the “original” meaning of the text?

In this paper, I argue that legal reason as traditionally employed within our constitutionally constrained hybrid common law and statutory legal system is not just one strategy among many (none with any better claim to legitimacy than the other) for resolving “hard” legal cases. Rather, it is a particular form of practical or moral reason that operates within the constraints of the positive law to keep

7. See DWORKIN, LAW’S EMPIRE, *supra* note 2, at 226.

8. See POSNER, PROBLEMATICS, *supra* note 6, at vii-viii.

9. See ANTONIN SCALIA, A MATTER OF INTERPRETATION: FEDERAL COURTS AND THE LAW 45 (1997).

10. DWORKIN, JUSTICE, *supra* note 3, at 211.

11. *Id.* at 211-12.

12. *Id.* at 212.

equity and law conjoined in the production of judgments that are fair to the parties, further the moral purpose of the law, maintain the integrity and functionality of the law, and ensure justice. To analyze traditional legal reason and justify its use in *every* case is thus the objective of this paper.

II. JUDICIAL STRATEGY AND LEGAL REASON

The justification for judges employing alternative legal strategies to resolve hard, often constitutional, cases depends entirely upon the validity of the anti-traditionalist argument—the argument that traditional jurisprudence, grounded in the interpretation of rules and principles in the positive law as applied in particular cases and controversies, is not up to the task of solving “hard cases” in developing and divisive areas of law. So how does that argument describe traditional jurisprudence and its shortcomings? Is that description accurate? And what do contemporary legal philosophers propose to put in its place to solve the problems they contend it cannot solve?

Dworkin states that traditional jurisprudence, or “[c]onventionalism[,] requires judges to study law reports . . . to discover what decisions have been made by institutions conventionally recognized to have legislative power” and to interpret legal text accordingly.¹³ He then argues that judicial opinions based solely on those “conventional” sources in the positive law are merely “backward-looking factual reports” that cannot resolve the novel and controversial moral issues presented by legal cases.¹⁴ He states,

Law by convention is never complete, because new issues constantly arise that have not been settled one way or the other by whatever institutions have conventional authority to decide them. So conventionalists add this proviso to their account of legal practice. “Judges must decide such novel cases as best they can, but by hypothesis no party has any right to win flowing from past collective decisions—no party has a *legal* right to win—because the only rights of that character are those established by convention. So the decision a judge must make in hard cases is discretionary in this strong sense: it is left open by the correct understanding of past decisions. A judge must find some other kind of justification beyond law’s warrant, beyond any requirement of consistency with decisions made in the past, to support what he then does. (This might lie in abstract justice, or in the general interest, or in some other forward-looking justification.) Of course convention may convert novel decisions into legal rights for the future. . . . In this way the system of rules sanctioned by convention grows steadily in our legal

13. See DWORKIN, LAW’S EMPIRE, *supra* note 2, at 226. For legal positivist theory, see H.L.A. HART, THE CONCEPT OF LAW 89-95 (1961); *see also* DWORKIN, LAW’S EMPIRE, *supra* note 2, at 114-17, 430-33 (setting out his conception of conventionalism and of Hart’s legal positivism); Adam Liptak, *The Transcendent Lawyer*, 2005 LAW SCH. 14, 15-16, available at http://issuu.com/nyulaw/docs/2005?mode=9_p.

14. DWORKIN, LAW’S EMPIRE, *supra* note 2, at 225.

practice.”¹⁵

Dworkin advocates an aspirational or *perfectionist* theory of legal reason to resolve the cases he contends conventional jurisprudence cannot solve.¹⁶ He argues that American legal principles, at least at the constitutional level, state universal moral truths and that the duty of judges of integrity is to discern and apply those truths in deciding “hard cases.”¹⁷ These moral truths are universal in form and objective—both metaphysically, in that there exist right answers to legal questions, and epistemically, in that there are mechanisms for discovering right answers free of distorting factors.¹⁸ Once these are discovered, objectively true propositions of law may be deduced from them and a just society implemented.¹⁹ Thus, the main inquiry for perfectionist legal philosophers and judges is what conditions must hold so that legal judgments may follow deductively from objectively true propositions of law.²⁰ Dworkin states, “This is particularly important in political communities like our own in which important political decisions are made by judges who are thought to have a responsibility to decide only as required or licensed by true propositions of law.”²¹

Richard Posner, generally Dworkin’s antagonist, agrees with Dworkin’s characterization of traditional legal reasoning as “conventional” and unable to solve “hard cases.”²² But he opposes it with his own theory of judicial *pragmatism*. This theory holds that the ultimate aim of society in legal

15. *Id.* at 115 (citation omitted).

16. See, e.g., DWORKIN, JUSTICE, *supra* note 3, at 2, 5, 187 (describing his own theory as doctrinal and aspirational and setting out the anti-traditionalist argument); DWORKIN, LAW’S EMPIRE, *supra* note 2, at 114-15, 225; RONALD DWORKIN, TAKING RIGHTS SERIOUSLY 14, 82-84 (1978) [hereinafter DWORKIN, RIGHTS]; see also CASS R. SUNSTEIN, RADICALS IN ROBES: WHY EXTREME RIGHT-WING COURTS ARE WRONG FOR AMERICA 31-39 (2005) [hereinafter SUNSTEIN, RADICALS] (using the term “perfectionist” for jurisprudential theories like Dworkin’s).

17. See, e.g., RONALD DWORKIN, FREEDOM’S LAW: THE MORAL READING OF THE AMERICAN CONSTITUTION 2-13 (1996) [hereinafter DWORKIN, FREEDOM’S LAW] (describing the “moral reading” of the Constitution); DWORKIN, LAW’S EMPIRE, *supra* note 2, at 225-28.

18. See Brian Leiter, *Introduction*, in OBJECTIVITY IN LAW AND MORALS 1, 3 (Brian Leiter ed., 2001) [hereinafter OBJECTIVITY] (discussing objectivity in legal reasoning).

19. Rational idealist perfectionism dedicated to discovering and implementing the objectively true laws of the just society has been the dominant social and political theory in American philosophy for at least the last half century. See, e.g., DWORKIN, FREEDOM’S LAW, *supra* note 17, *passim*. For rational idealist theories of the just society, see generally THOMAS NAGEL, EQUALITY AND PARTIALITY (1991); JOHN RAWLS, A THEORY OF JUSTICE (1971); JOSEPH RAZ, THE MORALITY OF FREEDOM (1986).

20. See DWORKIN, JUSTICE, *supra* note 3, at 5 (“Our main question is about the nature of the doctrinal concept of law. We ask whether moral considerations figure among the truth conditions of propositions of law and, if so, how.”).

21. *Id.* at 2.

22. See POSNER, PROBLEMATICS, *supra* note 6, at ix-x.

decisionmaking, and of the judges who actualize society's aims, is "to maximize the social utility of law."²³ On this view,

the judge or other legal decisionmaker thrust into the open area, the area where the conventional sources of guidance run out (such sources as previously decided cases and clear statutory or constitutional texts), can do no better than to rely on notions of policy, common sense, personal and professional values, and intuition and opinion, including informed or crystallized public opinion.²⁴

The judge should be informed by "the analytic methods, empirical techniques, and findings of the social sciences (including history)."²⁵

Dworkin and Posner are not alone in denying the capacity of traditional legal reasoning within the positive law to resolve "hard cases" and in affirming the legitimacy of alternative judicial strategies. Cass Sunstein argues that there *is* no single valid form of sound legal reasoning; there are only judicial strategies.²⁶ Unlike the blindfolded image of justice balancing her scales, he claims, "judges have no scale . . . they must operate in the face of a particular kind of social heterogeneity: sharp and often intractable disagreements on matters of basic principle."²⁷ Thus, Sunstein also advocates taking a constructivist approach to the law to correct the defects of traditional jurisprudence—but only, he argues, where it really matters.

Sunstein agrees with Dworkin and Posner that legal rules are "approaches to law that aspire to make legal judgments in advance of actual cases,"²⁸ i.e., fixed principles that entail objectively true or false legal judgments. But unlike those theorists, Sunstein does not describe traditional judges as reasoning deductively from fixed conventional rules to the propositions of law entailed by them. Rather, he calls traditional legal reasoning "rulelessness" and "casuistry"—which he defines as "analysis of cases unaccompanied by rules."²⁹ Traditional jurisprudence, he argues, is properly described as ruleless precisely because its judgments do not derive from theory but are "case-by-case decisions, narrowly tailored to the particulars of individual circumstances."³⁰ He distinguishes "ruleless" traditional jurisprudence from theory-driven jurisprudence, or "rule-bound justice," which he describes as "an approach to law that specifies a simple and (usually) unitary value, that operates at a high level of abstraction, and that decides cases by bringing the general theory to bear," operating deductively so

23. *Id.* at xi.

24. *Id.* at viii.

25. *Id.*

26. CASS R. SUNSTEIN, *LEGAL REASONING AND POLITICAL CONFLICT*, at vii-x, 3-7 (1996) [hereinafter SUNSTEIN, *LEGAL REASONING*] (declaring the nonexistence of a single theory of legal reasoning).

27. *Id.* at 3.

28. *Id.* at 21 (emphasis omitted).

29. *Id.* at 10.

30. *Id.* at 11.

that “[r]esults in particular cases are viewed as a logical consequence of the general theory.”³¹

Sunstein identifies fundamentalists as contemporary theory-driven constitutional strategists or originalists, for whom the “goal is to return to what they see as the essential source of constitutional meaning: the views of those who ratified the document,”³² and perfectionists, who “believe that the continuing judicial task is to make the . . . [Constitution] as good as it can be by interpreting its broad terms in a way that casts its ideals in the best possible light.”³³ Democratic perfectionists, he states, “believe that where the Constitution is ambiguous, judges should interpret it to promote democracy rather than to compromise it,” and they insist that the Supreme Court “act most aggressively when the requirements of democracy are themselves at risk,” believing “that the Court should protect those groups that are least able to protect themselves in democratic arenas.”³⁴ Rights perfectionists “insist that the Constitution should be read to protect the essentials of human dignity, including a right to make the most fundamental choices free from the constraining arm of the government.”³⁵

Observing that “[g]eneral theories are a natural ally of codification, which tries to organize and systematize the law, and a natural enemy of the common law, which tends to be quite unruly and to resist explanation according to theory,”³⁶ Sunstein posits a continuum with theory-driven jurisprudence at one end and “rulelessness,” or traditional common-law jurisprudence, at the other. Given the intractability of the theories of jurisprudence he identifies as being at the high end of the continuum and the unruly common-law system in which they are brought to bear at the other, Sunstein proposes that a “well-functioning legal system” might “adopt a special strategy for producing stability and agreement in the midst of social disagreement and pluralism”—namely, “*incompletely theorized agreements*”³⁷ in which “judges accept a certain approach to free speech, or equality, or religious freedom without necessarily agreeing on the deepest foundations of that approach.”³⁸ Judges can then resolve the dilemma posed by profound disagreements of principle at the high end of the continuum by moving from abstraction to a level of greater particularity, concentrating on

31. *Id.* at 14.

32. SUNSTEIN, RADICALS, *supra* note 16, at 26; *see generally* Frank B. Cross, *Decisionmaking in the U.S. Circuit Courts of Appeals*, 91 CAL. L. REV. 1457 (2003) (distinguishing four types of judicial decisionmaking: legal, political, strategic, and litigant-driven); Timothy P. Terrell, *Statutory Epistemology: Mapping the Interpretation Debate*, 53 EMORY L.J. 523 (2004) (setting out different theories of judicial interpretation).

33. SUNSTEIN, RADICALS, *supra* note 16, at 32.

34. *Id.* at 38-39.

35. *Id.* at 39.

36. SUNSTEIN, LEGAL REASONING, *supra* note 26, at 16. Of course, part of the objective of this paper is to explain our traditional “quite unruly” common law jurisprudence!

37. *Id.* at 4 (emphasis in original).

38. Cass R. Sunstein, *The Minimalist Constitution*, in THE CONSTITUTION IN 2020, at 37, 41 (Jack M. Balkin & Reva B. Siegel eds., 2009).

the lower-level principles on which they agree, “understanding or converging on an ultimate ground for that acceptance,”³⁹ and being “cautious about undoing the fabric of existing law.”⁴⁰ He calls his theory *minimalism* in that it is intended to minimize harm done to the legal system when judges resort to theory to decide “hard cases” by confining theory-driven decisionmaking only to the most important cases of principle.⁴¹

William Eskridge, unlike other contemporary legal philosophers, approaches the subject of sound judging through a comprehensive critical survey of all leading current methodologies of statutory interpretation.⁴² His argument is, however, in my view, as much applicable to theories of legal reason and legal strategy in general as it is simply to theories of statutory construction. Rather than arguing as an advocate that judges of integrity derive their judgments from objective moral principles, a conception of social utility, fixed original texts, or conventionally decreed mandates, Eskridge steps back and asks what statutory interpretation—I would say, more broadly, legal interpretation or even legal reason—actually is. He argues that it is an ongoing, “dynamic” process and that “[t]he work of interpretation is to *concretize* the law in each specific case—i.e., it is the work of *application*.⁴³ Thus, he seeks a determinate methodology of interpretation that can make legal decisions concrete and predictable within a dynamic social process. However, upon analysis, he concludes that each of the current interpretative legal theories is incapable of accommodating both dynamism and the determinacy required to concretize the application of the law.⁴⁴

Eskridge rejects foundationalist theories of legal interpretation as unable to “yield analytically determinate answers in the hard cases.”⁴⁵ Originalism is unsatisfactory because “[t]he ‘original intent’ and ‘plain meaning’ rhetoric in American statutory interpretation scholarship and decisions treats statutes as static texts and assumes that the meaning of a statute is fixed from the date of enactment.”⁴⁶ Similarly, classical liberalism “views government as a social contract among autonomous individuals who in the distant hypothetical past gave up some of their freedom to escape the difficulties inherent in the state of nature.

39. SUNSTEIN, *LEGAL REASONING*, *supra* note 26, at 5.

40. SUNSTEIN, *RADICALS*, *supra* note 16, at 29.

41. *See id.* at 28-29.

42. WILLIAM N. ESKRIDGE, JR., *DYNAMIC STATUTORY INTERPRETATION* 14 (1994).

43. *Id.* at 64 (quoting HANS-GEORG GADAMER, *TRUTH AND METHOD* 329 (J. Weinsheimer & D. Marshall trans., 2d rev. English ed. 1991)).

44. *See id.* at 133.

45. *Id.*; *see also id.* at 47 (noting limitations of originalism and observing that interpreters are interested in text and the intention of the enacting body, but also in “the facts and equities of the case, precedents interpreting the statute and legislative feedback, and the consequences of accepting one interpretation over another”); *id.* at 14 (arguing that no foundationalist theory of statutory interpretation (purposivism, intentionalism, textualism) “yields determinate results” or “fully constrains statutory interpreters or limits them to the preferences of the enacting coalition”).

46. *Id.* at 9.

. . . Hence, laws enacted by prescribed constitutional procedures are legitimated by the consent expressed in this original social contract.”⁴⁷ But, Eskridge argues,

[s]uch indeterminacy [as that entailed by a theory that looks back to the eighteenth century to determine the fixed meaning of contemporary legal terms] creates a dilemma for liberal theories of statutory interpretation. If important policy issues cannot be resolved by an interpretive methodology that is determinate and predictable, then unelected judges and agencies will make political decisions that liberalism leaves to the legislature or some other majoritarian institution.⁴⁸

Thus, he rejects classical liberalism as well.

Eskridge applauds the dynamism of the *legal process theory* of interpretation,⁴⁹ which interprets statutes (or constitutional principles) by rationally deducing the answer to divisive legal questions from the best constructive social policy. He states:

Whereas liberal theory posits mutually suspicious humans who form a social contract to escape the state of nature, legal process theory posits humans who recognize their interdependence and cooperate for the advancement of common interests. The state exists to further the interests that the members of a community have in common. Legal process views law as a “purposive activity, a continuous striving to solve the basic problems of social living.”

Legal process theory invites dynamic statutory interpretation. Viewing statutes in rationalist terms, most legal process thinkers accept dynamic interpretation as normatively essential to the implementation of statutory policy.⁵⁰

However, he ultimately rejects legal process theory too, pointing out:

The main difficulty with the rationalist tradition as applied by legal process writers is summed up by the question: Whose reason? Rather than identifying right answers in the hard cases, the application of law’s reason depends on the interpreter’s own policy choices, which are themselves guided by the framework she brings to the issue.⁵¹

Eskridge himself tentatively offers as a potential model of sound legal interpretation a hermeneutical model of statutory interpretation derived from the method of deconstructing literary texts developed by Hans-Georg Gadamer and

47. *Id.* at 111.

48. *Id.* at 133 (internal citation omitted).

49. *Id.* at 11, 141-42; see generally HENRY M. HART, JR. & ALBERT M. SACKS, THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW (1994) (setting out the theory).

50. ESKRIDGE, *supra* note 42, at 141-42 (internal citation omitted).

51. *Id.* at 142.

others.⁵² This method recognizes the “limits imposed by tradition” and the surrounding and professional culture on the subject matter being considered—here, the law.⁵³ While Eskridge acknowledges the distinctions between literary and legal interpretation, he argues that the

lessons of legal hermeneutics—the importance of the interpreter’s horizon, the evolving nature of the text’s horizon, and the importance of application to new factual circumstances—make possible the construction of a model of statutory interpretation that . . . recasts the traditional textual, historical, and evolutive inquiries as more explicitly interconnected and mutually influencing.⁵⁴

The problem with the hermeneutical method, however, in Eskridge’s view, is that it fails to provide concretizing constraints on sound judicial decisionmaking. Thus, to correct hermeneutics’ defects, he opts for what he calls a “*critical pragmatism*”⁵⁵ that incorporates elements of each of the theories he has analyzed. He takes this “pragmatist approach” to interpretation because of his belief that “the interpreter’s fidelity to the rule of law is nothing more . . . than a sympathetic effort to understand a statute in the context of the problem at hand and of ongoing practice.”⁵⁶ He then grafts onto that approach a “critical” stance that would permit “an ongoing reevaluation of practice, especially dividing practices that marginalize groups of citizens and interest group distributions at the expense of the public interest.”⁵⁷ Eskridge acknowledges, however, that a critical theory of interpretation assigns to judges precisely the policymaking role that he himself recognizes as a critical flaw in both foundationalist and legal process theories of legal interpretation.⁵⁸ As he puts it, “social constructionism undermines democratic theory as a legitimating device.”⁵⁹ Thus, he acknowledges that ultimately even his own critical pragmatism fails to solve the problem of providing a theory of legal analysis or legal reason that is at once dynamic and concretizing and also compatible with democratic theory.⁶⁰

The one theory of legal reason and analysis that Eskridge does *not* analyze is traditional jurisprudence, presumably because, as Dworkin points out, “it is no longer an important force either in legal practice or in legal education.”⁶¹ Thus, he does not acknowledge that there is already present in the jurisprudential theater a methodology that gives *legally* constrained, although not fully

52. *Id.* at 61-68, 348 n.33.

53. *Id.* at 65.

54. *Id.* at 63.

55. *Id.* at 193 (emphasis in original).

56. *Id.*

57. *Id.*

58. *See id.*

59. *Id.*

60. *See id.*

61. DWORKIN, JUSTICE, *supra* note 3, at 211.

determinate, answers to legal questions—a methodology that both accommodates the dynamism of the law and society and concretizes the law in particular cases while maintaining the purpose, integrity, and functionality of the law over time. The argument for that overlooked methodology is made below.

II. PRACTICAL REASON⁶²

I agree with Eskridge that statutory interpretation—and legal reasoning, or the solving of legal problems—is a dynamic process and that “[t]he work of interpretation is to *concretize* the law in each specific case—i.e., it is the work of *application*.⁶³ I further agree that the “lessons of legal hermeneutics—the importance of the interpreter’s horizon, the evolving nature of the text’s horizon, and the importance of application to new factual circumstances—make possible the construction of a model of statutory interpretation that . . . recasts the traditional textual, historical, and evolutive inquiries as more explicitly interconnected and mutually influencing.”⁶⁴ The question is: Of what does a satisfactory dynamic and concretizing model of statutory interpretation—or, more generally, a satisfactory model of reasoning to a valid and sound legal judgment, consist?

Most leading contemporary legal scholars agree on this, at least: that the methodology that judges of integrity should follow is theory-driven and rationally realizes the ends proposed by the theory, whether those ends be the judge’s reconstruction of the original intentions of the drafters of legislation or of the Constitution, or the instantiation of “true” moral propositions, or the maximization of social utility within the context of a case. But, I shall argue, the effectively organic, empirical nature of legal problem solving resists the rationalist shoe-horning of the law into abstract theoretical boxes. However, this is precisely because traditional legal reasoning accommodates both empiricism and rationality, concretizing abstract legal concepts in particular cases within a dynamic system of empirical laws. Thus, it is not merely one among many strategies for producing valid and sound legal judgments. Rather, it is a particular method of reasoning from abstract legal principles to sound prescriptions for action under empirical conditions that both accommodates the dynamism of the law and society and concretizes the law in particular cases, all while maintaining the purpose, integrity, and functionality of the positive law over time. What, then, is legal reason on the traditional jurisprudential model?

62. The argument that follows in this section is essentially a recapitulation of arguments I have made in prior articles. See Evelyn Keyes, *The Just Society and the Liberal State: Classical and Contemporary Liberalism and the Problem of Consent*, 9 GEO. J.L. & PUB. POL’Y 1 (2011) [hereinafter Keyes, *The Just Society and the Liberal State*]; Evelyn Keyes, *Two Conceptions of Judicial Integrity: Traditional and Perfectionist Approaches to Issues of Morality and Social Justice*, 22 NOTRE DAME J.L. ETHICS & PUB. POL’Y 233 (2008) [hereinafter Keyes, *Two Conceptions of Judicial Integrity*].

63. ESKRIDGE, *supra* note 42, at 64 (quoting GADAMER, *supra* note 43, at 329).

64. *Id.* at 63.

I take it as given that legal reason has a peculiar subject matter (the law) in which systemically valid conclusions, or judgments, follow rationally from the application of legal principles to the facts of particular cases and prescribe specific actions to be taken. More than one valid rational argument can be made from any abstract legal concept, principle, or rule whose meaning is not concretely fixed, however. Therefore, legal reason requires both a method for interpreting concepts and a means of reconciling conflicting interpretations and determining which is better and should be instantiated. Thus, the process of legal reason is necessarily more than the mere logical deduction of valid legal conclusions or judgments from objectively defined principles such as Dworkin's "true" moral principles of liberty and equality, Posner's principles of social utility, Scalia's original constitutional text, or even fixed "conventional" principles.

Legal reason is, indeed, a rational process, but it is also a practical, interpretive, and evaluative process that takes place within a constraining system of pre-existing laws. By its means, abstract value-laden legal concepts are applied and reconciled, and legal judgments are issued and implemented under empirical circumstances to tell us what we *must* or *should* do. Legal reason is thus not only a rational process, but also a normative process, or a process for implementing standards of value through prescriptions for action.

A. The Categorical Properties of Moral Reason

That legal reason applies value-laden principles and instantiates value-laden rights to reach prescriptive judgments entails that it is a particular type of applied moral reason,⁶⁵ or practical reason,⁶⁶ as opposed to purely deductive reason.

65. This assertion presupposes that there is an integral relationship between morality and law, and part of the purpose of this paper is to justify that presupposition. In that regard, I note that although the concept of legal reason as a type of applied moral reason that is developed in this paper is mine, the inquiry into the relationship between morality and the positive law is not new.

Hart, for example, developed legal positivist theory without defining morality; he avers that there are many different types of relations between law and morality, but "nothing which can be profitably singled out for study as *the* relation between them." HART, *supra* note 13, at 181. Starting, therefore, with rules consisting of "certain basic protections and freedoms" that already contain the concepts of both morality and law, he adopts a conventionalist concept of law as consisting of "all rules which are valid by the formal tests of a system of primary and secondary rules, even though some of them offend against a society's own morality or against what we may hold to be an enlightened or true morality." *Id.* at 201, 205. This concept, he contends, has the virtue of allowing "the invalidity of law to be distinguished from its immorality . . . [enabling] us to see the complexity and variety of these separate issues." *Id.* at 207. Hart acknowledges, however, that the moral contention "that a legal system must treat all human beings within its scope as entitled to certain basic protections and freedoms, is now generally accepted as . . . an ideal of obvious relevance in the criticism of law." *Id.* at 201. He concedes that "[i]t may even be the case that a morality which does not take this view of the right of all men to equal consideration, can be shown by philosophy to be involved in some inner contradiction, dogmatism, or irrationality," in

Therefore, I begin the inquiry into the nature of legal reason with the question, “What is the nature of applied moral reason or practical reason?” For only when we have determined what practical reason is, and how it functions to preserve moral value within an empirical context, can we determine how legal reason, as a type of practical reason, functions to preserve the values in the positive law.⁶⁷

First, I take it as axiomatic that the concept of morality itself has no meaning or extension without the concept of moral value. Likewise, the concept of moral value has no meaning except by reference to life, i.e., life alone has intrinsic moral value or worth. If so, then all human beings, simply by virtue of their humanity, have dignity, or intrinsic worthiness to be treated with respect.⁶⁸ Thus, the concept of the intrinsic worth of every person lies at the core of morality. All moral values other than the value of life itself are derivatively values by virtue of their qualities of respecting the worth of every person and sustaining and advancing life and its prospects. Therefore, I define “morality” categorically as the complex of principles, rules, rights, obligations, judgments, and prescriptions that derive from and entail recognition of and respect for the intrinsic value of

which case “the enlightened morality which recognizes these rights has special credentials as the true morality, and is not just one among many possible moralities.” *Id.* In other words, Hart does not deny the possibility that morality and the positive law may be shown to be more intimately connected than he recognizes.

Dworkin, reacting to Hart, argues that the essential problem with the positive law is its disjunctive “true” propositions of law, which are those derivable from the philosophically best construction of the moral principles incorporated into the United States Constitution. *See supra* text accompanying notes 16-21.

Sunstein, by contrast, takes the position that just as there is no single concept of legal reason, so there is no definition of morality. Rather, “[t]o understand what morality requires, or what the law should be in hard cases, we need to canvass what we—each of us—actually believe; there is no other place to look.” SUNSTEIN, *LEGAL REASONING*, *supra* note 26, at 18.

66. The term “practical reason” as a descriptive term for the type of reason used in applying moral principles to guide behavior is traceable in modern moral and political philosophy to Kant’s definition of practical reason in his *Foundation of the Metaphysics of Morals*:

Everything in nature works according to laws. Only a rational being has the capacity of acting according to the conception of laws, i.e., according to principles. This capacity is will. Since reason is required for the derivation of actions from laws, will is nothing else than practical reason. . . . That is, the will is a faculty of choosing only that which reason, independently of inclination, recognizes as practically necessary, i.e., as good.

IMMANUEL KANT, FOUNDATIONS OF THE METAPHYSICS OF MORALS AND WHAT IS ENLIGHTENMENT? 29 (Lewis White Beck trans., Bobbs-Merrill Co. 1959) (1785).

67. If this argument seems too abstruse, I invite the reader to skip this section and take up the argument in the next section, where the building upon the foundations begins.

68. See KANT, *supra* note 66, at 9, 46-49. This conception of core value reflects the Kantian conception of a person “as an end in himself.” *Id.* at 47. Kant, however, located intrinsic value in the pure rational will rather than in life itself. *See id.* at 9.

human life⁶⁹ and that are directed to the furtherance of those projects which enhance the quality of our lives and contribute to their flourishing—that contribute to the *good* as empirically realized.⁷⁰ And I define applied moral reason, or practical reason, as reason directed to making applied, categorically moral, judgments.

It follows from the definition of morality and moral reason that people are “moral” (and societies “just”) precisely insofar as they inculcate and adhere to principles, rules, and methods of reasoning and making judgments that recognize and respect the intrinsic dignity and worth of every person and that make and implement moral judgments rationally in accordance with personal or social codes (or systems of principles and rules) that advance the flourishing lives of those affected. But if people and societies are moral only insofar as personal, social, or societal guiding principles, rules, and judgments are intrinsically moral, and if the positive law of a society consists of those general rules the society makes and enforces through its official governmental institutions, then a society is moral and its laws just only if its positive law and its means of making and interpreting the law respect the intrinsic dignity and worth of each member of society and further the flourishing life of the whole society—the *common good*. But how can we rationally get from the concept of morality to the role of applied moral reason, or practical reason, as the guarantor of the morality of empirical moral judgments, and from there to the concept of the positive law as an intrinsically moral social system and to the role of legal reason as the guarantor of social justice within that system? To answer these questions, we must first ask what the concept of morality entails that makes a judgment or prescription moral.

If morality is understood as respect for the intrinsic dignity and worth of every human being, then all human beings have not only intrinsic moral worth but also moral interests, i.e., the intrinsic right to be treated with respect, both

69. This view is shared by Dworkin, who has stated that “[w]e almost all accept . . . that human life in all its forms is *sacred*—that it has intrinsic and objective value quite apart from any value it might have to the person whose life it is” and who likewise takes “the abstract right to concern and respect . . . to be fundamental and axiomatic.” DWORKIN, RIGHTS, *supra* note 16, at xiv-xv; Ronald Dworkin, *Life is Sacred. That's the Easy Part*, N.Y. TIMES, May 16, 1993, at 36; see also DWORKIN, FREEDOM'S LAW, *supra* note 17, at 84.

70. I accept Aristotle's argument that the objective of morality is the achievement of eudaimonia, defined not simply as the equally abstract term “happiness,” but as a flourishing life. See ARISTOTLE, NICOMACHEAN ETHICS 5 (Terence Irwin trans., 1985) (“[M]ost people virtually agree [about what the good is], since both the many and the cultivated call it happiness [eudaimonia], and suppose that living well and doing well are the same as being happy. But they disagree about what happiness is. . . .” (first alteration in original)); see also *id.* at 19 (“[T]he end is a sort of living well and doing well in action”); cf. KANT, *supra* note 66, at 35 (distinguishing morality as categorically a creature of pure reason from the maximization of happiness as an empirical goal and stating, “[A]ll elements which belong to the concept of happiness are empirical, i.e., they must be taken from experience, while for the idea of happiness an absolute whole, a maximum, of well-being is needed in my present and in every future condition.”).

substantively and procedurally, regarding matters that affect their lives.⁷¹ Moreover, because there is nothing in the concept of morality itself to distinguish a moral interest in one person from the same interest in another person, the concept of morality logically requires that all equal moral interests be treated equally. In other words, the moral interests of any person affected by a moral choice should be treated as equal to the same interests of every other similarly situated person. Morality, as respect for the intrinsic dignity and worth of every person, also logically requires that all moral agents respect the right of each rational person as a moral agent to determine individually how the dignity and worth of each person affected by his decision would be best respected as if he himself were in any of the positions subject to his own decision and thus were responsible for his own well-being. In other words, morality as respect for the equal individual dignity and worth of every person entails respect not only for the *equality* of all persons in any given position with respect to a moral choice, but also for the autonomy, or *liberty*, of each moral agent to perceive himself and others as potential objects of moral judgments and to act accordingly. To ensure the morality of his prescriptions, therefore, each moral agent must make judgments as if he himself were subject to them and had the same respect for the moral interests, or *rights*, of those persons potentially affected by his prescriptions as he would have for his own if he were in their positions.⁷² To put these Kantian concepts in Kantian terms, “[t]his principle of humanity and of every rational creature as an end in itself is the supreme limiting condition on freedom of the [rational moral] actions of each man.”⁷³

The concept of rational moral decisionmaking thus requires us to recognize that all persons have intrinsic moral worth. Additionally, all persons have moral interests correlated with abstract moral rights and obligations equally inhering in all similarly situated persons that moral agents may rationally recognize and instantiate. These rights and obligations, in turn, correlate with moral principles generally applicable to all similarly situated persons. For example, the moral principle “stealing is wrong”—in its imperative form, “Thou shalt not steal,” and in its judgmental and prescriptive form, “Thou hast wrongfully stolen from thy neighbor and therefore must redress the wrong”—correlates with a personal right not to be stolen from, a personal obligation not to steal, and an obligation of the moral agent to enforce this principle equally for himself and others in the same position.

Moreover, because all rational persons are potentially moral agents, or makers of moral judgments, as well as objects of moral decisionmaking, and because all rational persons are equally entitled to respect in both roles, the general law must be such that a moral agent reasoning morally would willingly

71. Conversely, treating human beings as lacking intrinsic worth—treating them as mere objects or commodities without moral interests intrinsically worthy of respect—is categorically inconsistent with morality and with law. The insight is, of course, Kant’s. See KANT, *supra* note 66, at 9, 46-49, 51.

72. Cf. *id.* at 51-52 (describing moral agents as universal legislators in a world of ends).

73. *Id.* at 49.

subject himself to it. Therefore, not only does a moral agent not steal personally, he also cannot morally prescribe that some persons may steal while others similarly situated may not or that some persons may be stolen from while others similarly situated may not be. Moral prescriptions are impartial as to all persons and, therefore, *universalizable*.

These meta-principles of moral decisionmaking are categorical; thus, they apply universally, or in every case of valid moral reasoning. In Kantian terms, they collectively constitute a categorical imperative of rational morality or practical reason.⁷⁴ Indeed, we can say that moral judgments and prescriptions made in accordance with the rules of practical reason are procedurally fair precisely because they respect the liberty and equality of all persons both as agents and as objects of moral decisionmaking under impartial, universalizable principles.

However, the formal constraints placed on moral reasoning by universal moral principles of liberty and equality do not tell us how a moral agent can rationally determine whether employing an intrinsically rational and fair decisionmaking process will have good empirical consequences. That is, the categorical requirements of moral reason fail to tell us how to determine what practical judgments among those available would best advance the flourishing life of those affected by them or the good as actually realized. What, then, ensures not only that applied moral judgments are procedurally fair, but also that they are best suited to advance their purpose under actual empirical conditions—or that they are sound?

B. The Practical Application of Moral Reason

Moral judgments, by definition, entail prescriptions for action under empirical conditions. Thus, they necessarily have empirical content, follow from empirical principles applicable under empirical circumstances, and instantiate empirical rights and obligations of persons in particular conceptual and factual positions, affecting real people under real circumstances. Essential to the understanding of practical moral judgment, therefore, is the recognition that the empirical conditions under which applied moral judgments are made impose their own constraints on moral evaluation that complement the formal constraints of moral reason. Thus, applied moral reasoning is necessarily a more complex process than mere rational deduction of valid conclusions from intrinsically moral principles. What, then, are the empirical constraints upon practical reason, or applied moral reason, and how do they impact our moral judgments?

First, despite their abstract universal form, no empirical moral principles are applicable in all circumstances. Rather, empirical moral principles or imperatives—such as the affirmative command to respect one's parents and the

74. Cf. *id.* at 38-39 (stating that since the categorical imperative “contains beside the law only the necessity that the maxim should accord with this law, while the law contains no condition to which it is restricted, there is nothing remaining in it except the universality of law as such to which the maxim of the action should conform” (internal citation omitted)).

negative commands not to steal or commit murder—although objective, universal, and categorically moral in form, are hypothetical in application. That is, they are substantive in content, limited in practical scope, and purposive or instrumental in effect.⁷⁵ They apply only in certain circumstances, and only those moral principles that have a correlative relation with the interests of persons who will be affected by a particular moral judgment are implicated in the process of moral reasoning.

Second, the interests of persons in different empirical positions are not the same and generate different moral judgments with respect to each position. The killer who sprays a room with bullets is in a different position than the killed, both logically and morally, with respect to that act. Likewise, our moral judgments regarding persons in the same positions, such as that of the killer, differ in varying factual circumstances. Suppose violent terrorists aim assault rifles at undercover security guards who shoot them first. Suppose the reverse. Suppose terrorists shoot children held hostage in a school. Suppose an angry, estranged, or distraught husband kills his wife and children. Suppose the killer and the killed are drug addicts and the cause of the shooting is a deal gone sour. Our moral views about the killers are different in each instance. Empirical moral judgments are shaped by empirical facts, by precepts learned, by background knowledge, and by perceptions of the circumstances. They are shaped by conceptual and factual context.

Third, practical moral judgments are made under imperfect conditions of knowledge as to the material facts and the consequences of the judgment. Thus, the resolution of empirical moral problems is confined to material considerations and foreseeable consequences.

Finally, the moral interests of persons affected by a moral judgment may, and often do, conflict with each other and with other interests, generating an empirical moral nexus that requires a method of dispute resolution in which only one empirical principle or rule, or only one interpretation of a principle or rule, can emerge as the rule of prescription, and only one interest or right may be instantiated by the moral agent at the expense of another. Thus, practical moral judgment requires an interpretive and evaluative process for resolving conflicts, using as its operational tools rules of issue identification, interpretation and construction, and standards of evaluation, as well as deductive logic. These tools of what we may call the moral analytic allow a rational and fair-minded moral agent to define and weigh empirical alternatives and to resolve conflicts among the potential outcomes of his choice by deciding which of two or more interpretations of a concept is most consistent with the facts and the body of relevant moral concepts and which among alternative prescriptions best advances the moral interests of those affected, consistent with fairness and the maintenance of the purpose, integrity, and functionality of the system as a whole, and thus

75. The term is again Kant's. Cf. *id.* at 31 (distinguishing hypothetical imperatives, which are limited in scope and directed to some purpose or “good only as a means to something else,” from the categorical imperative, which is thought of as good in itself and “hence as necessary in a will which of itself conforms to reason as the principle of this will”).

should be implemented.

Practical reason is thus an ongoing and indeterminate process of concretizing moral precepts in moral judgments under empirical conditions through repeated applications of moral principles, rules, and standards; comparing the consequences of different interpretations of principles; and instantiating rights and obligations in different conceptual and factual contexts to achieve a moral purpose within the formal constraints of fairness imposed by the categorical moral imperative. The process of reasoning morally prescribes not only which interpretations are best and which rights are weightiest in particular circumstances and should be implemented, but also which rules and standards are best to secure a functional and moral system. In other words, the interpretive and evaluative process itself prescribes the empirical scope of principles and the relative weight of rights in different types of circumstances and thus determines which moral judgment is not only rational and fair, but best in any given case or controversy.

Over time the process of repeatedly making rational practical moral judgments creates an evolving hierarchy of more or less well-ordered and well-defined principles and rules, with moral principles being abstract, universal concepts; moral rules being subordinate, more particularized subsets of principles; rules of interpretation and construction being operations for determining the boundaries of those sets of factual circumstances that fall within the scope of a legal rule; and standards of evaluation being operations for evaluating outcomes and determining which is best. Within this hierarchy, fundamental rights are those moral interests that dominate and determine the outcome of empirical judgments at the most abstract and comprehensive level in case after case because their instantiation furthers ends rational moral agents subject to the system deem essential to the preservation and furtherance of the flourishing life through a system of ordered moral liberties. These are, therefore, most general in their application and least subject to radical change.

Within an applied moral system, each judgment or prescription is concrete and fixed with respect to those interests directly affected, but it enters into and becomes integral with a non-determinate, dynamic, and open-ended set of heteronomous moral principles and their subordinate rules. These conform to the categorical requirements of morality and are used to make rational moral judgments and prescriptions when logically applied to particular sets of facts under applicable evaluative standards. The judgments the process of moral reasoning produces within such a self-creating, self-correcting system then become part of and help define the system of flexible, open-ended rules and principles that constitute the moral system itself.

In such a system, a moral agent faced with making an empirical moral judgment or prescription takes into account the operative procedural and substantive principles and rules within the system, the facts of the case, and the relative weight of those principles and correlative rights whose instantiation is sought under applicable evaluative standards; he interprets moral concepts and evaluates alternatives for each position impartially as if he himself were to be affected by his judgment, ensuring his responsibility as a moral agent. When the substantive principles followed and the rights instantiated are intrinsically fair

and rationally judged most conducive to the good by the agent, the outcomes the process produces—its judgments and prescriptions—will be themselves intrinsically rational and moral, and therefore just, and they will contribute to the integrity and functionality of the moral code into which they enter. There is no need—and, indeed, no justification—for a moral agent to seek outside the system of applied moral principles the “true” principles that govern the “best” constructive interpretation of morality or the “best” consequences of his decision in terms of social utility.

An empirical moral system is not derived from a simple set of objectively true moral principles or guidelines, regardless of the contentions of rationalist moral philosophy. Rather, an intrinsically moral empirical value system is grounded in abstract rules and principles tested and incrementally adjusted over a long period of time in a variety of empirical situations and ultimately regularized in formal and informal personal, organizational, and societal moral codes. The system is kept fair and its empirical results just precisely insofar as its agents employ practical reason to reach and implement particular, concretizing moral judgments that instantiate moral rights correlated with moral principles and their subordinate rules, while maintaining the moral purpose, functionality, and integrity of the system.⁷⁶ But the system does not thereby become determinate and fixed. Rather, the actualization of morality remains a self-creating, bounded but open-ended, indeterminate, and dynamic process of concretizing or instantiating the principles of an ongoing moral system in particular cases under different circumstances through the exercise of moral reason and moral judgment. This process of reasoning morality creates a moral system consisting of moral prescriptions that, abstracted from, constitute moral rules that do justice to those affected by furthering their moral interests fairly to good ends while maintaining the integrity and functionality of the system itself over time.

But even if a moral system is a self-created body of moral principles and rules made in accordance with the formal requirements of practical reason, and if applied or practical moral judgments are integral parts of a dynamic empirical moral system, how can we justify the further claims that traditional legal reason is a form of practical reason and that it functions similarly to maintain the moral purpose, integrity, and functionality of a system of positive laws that is itself intrinsically moral?

III. LEGAL REASON

The principles that legal reason applies; the standards, rules, and precedents by which it interprets and applies them; the rights, obligations, and penalties a legal judgment instantiates; and the actions it prescribes are all part of a body of positive law self-created and self-perpetuated, like a moral system, through a

76. Correspondingly, if an empirical value system—whether personal or social—is not systematically moral, but incorporates principles and rules that conflict with the categorical requirements of morality, the moral integrity of the system can be maintained only by its agents’ rejection of the offending principles.

dynamic process actualized and implemented through concrete applications. In this process, legal judgments, like moral judgments, do not merely follow by deductive logic from the application of legal principles, standards, and rules under given circumstances, so that they are *valid*. They also require the interpretation of value-laden concepts and the evaluation of facts and alternatives under legal and evidentiary standards of evaluation and the rational and normative resolution of conflicts so that the judge may rationally reach prescriptions for action, or legal judgments, that are *sound*.

Within the system of laws in which legal reasoning takes place, no legal opinion or judgment stands alone, just as no moral judgment stands alone. Each relies upon past opinions and judgments, utilizes principles, rules, and standards in the law, and contributes to the ongoing, organic, self-creating, self-sustaining, and self-correcting—or autopoietic⁷⁷—system of publicly enforced legal principles, standards, and rights (or system of positive laws) in which it takes its place. And it is adherence to this complex process of legal reasoning within the context of the body of the law which ensures that a legal judgment will be accepted as both valid and sound and as conducive to the flourishing life of society—and therefore as just, by the parties and the public—and that it will contribute to the integrity and functionality of the dynamic and open-ended system of publicly enforceable laws of which it becomes a part. This dense thesis requires explication, however, and that is the objective of this section, which analyzes and justifies the concept of traditional legal reason as a complex type of practical reason employed within the context of the positive law.

It is self-evident that there is an integral relationship between personal and social moral codes. This relationship is grounded in the equal dignity and worth of every person. Further, it is inculcated and self-enforced in accordance with principles of practical reason on the one hand, and it is a social compact of self-made and self-enforced laws in the form of a constitutional representative democracy—founded, like our own, to ensure the safety and happiness of its members and grounded in principles of liberty, equality, and impartiality—on the other hand. Indeed, there is no serious question that American law was consciously erected on a moral base—namely, the original right of a free and equal people to establish for themselves a government of those laws they themselves deem most conducive to their own safety and happiness. The Declaration of Independence justified separation from the British crown and the formation of a new government on the quintessentially moral ground

that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and

77. The term “autopoiesis” is derived from the Greek word for “self-creating.” It was coined by biologists Humberto Maturana and Francisco Varela to describe living systems or autonomous, strictly bounded systems that are shaped by their interactions with the environment over time so as to maintain the system and the relations between parts. See HUMBERTO R. MATORANA & FRANCISCO J. VARELA, AUTOPOIESIS AND COGNITION: THE REALIZATION OF THE LIVING 78-79 (Robert S. Cohen & Marx W. Wartofsky eds., 1980).

the pursuit of Happiness. That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed, That whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their Safety and Happiness.⁷⁸

And the supreme instrument for structuring, protecting, and furthering that ideal was the American Constitution. Drafted by delegates of the people for their own governance and approved by the people through ratification by the states, the Constitution incorporates those intrinsically moral enabling principles, constraints on personal liberty, and exercises of governmental power that were, in the estimation of the Framers and the states that adopted the Constitution, essential to the concept of a just society of laws and grounded in those precepts deemed vital to the preservation and rational furtherance of the common good.⁷⁹

The Constitution expressly proclaims that the government of the United States was ordained and established by the representatives of the people “in Order to form a more perfect Union, establish Justice, insure domestic Tranquility, provide for the common defence, promote the general Welfare, and secure the Blessings of Liberty to ourselves and our Posterity.”⁸⁰ And in *Marbury v. Madison*, Chief Justice Marshall reaffirmed the “original right” of the

78. THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776).

79. See *McCulloch v. Maryland*, 17 U.S. (4 Wheat.) 316, 405-06 (1819) (“If any one proposition could command the universal assent of mankind, we might expect it would be this—that the government of the Union, though limited in its powers, is supreme within its sphere of action. This would seem to result, necessarily, from its nature. It is the government of all; its powers are delegated by all; it represents all, and acts for all. . . . But this question is not left to mere reason: the people have, in express terms, decided it, by saying, ‘this constitution, and the laws of the United States, which shall be made in pursuance thereof,’ ‘shall be the supreme law of the land’”); see also William J. Brennan, Jr., *The Constitution of the United States, in JUDGES ON JUDGING: VIEWS FROM THE BENCH*, *supra* note 4, at 188. Justice Brennan argues:

The Constitution on its face is, in large measure, a structuring text, a blueprint for government. . . . When one reflects on the text’s preoccupation with the scope of government as well as its shape, however, one comes to understand that what this text is about is the relationship of the individual and the state. The text marks the metes and bounds of official authority and individual autonomy. When one studies the boundary that the text marks, out, one gets a sense of the vision of the individual embodied in the Constitution.

As augmented by the Bill of Rights and the Civil War Amendments, this text is a sparkling vision of the supremacy of the human dignity of every individual. This vision is reflected in the very choice of democratic self-governance: the supreme value of a democracy is the presumed worth of each individual.

Id.

80. U.S. CONST. pmb1.

people to establish a government subject to those principles they themselves deem most conducive to their own collective happiness, stating, “That the people have an original right to establish, for their future government, such principles as, in their opinion, shall most conduce to their own happiness, is the basis, on which the whole American fabric has been erected.”⁸¹

As adopted, interpreted over time, and continually ratified by successive generations, the Constitution has remained the central structural document of the positive law, and hence of the social compact compounded of the positive laws. The Constitution not only creates the institutions of government and allocates delegated powers among them; it assures the people’s ultimate responsibility for ensuring that the government created by their ongoing delegation of authority to their representatives and ratified by their consent to its laws is morally and politically just.⁸²

In addition to the original moral and political right of self-government, which is a formal right referable to the categorical imperative itself, the Constitution sets out fundamental structural principles of ordered liberty—or enabling principles and constraints upon the collective will—regulating the body politic in an orderly, state-enforced, and intrinsically moral manner. And through the Bill of Rights and subsequent amendments, it ensures the fundamental individual rights of the American people against governmental or private infringement.

Substantive individual moral rights enumerated in the Constitution (and hence in the positive law) and held by the people against the State include, *inter alia*, the rights of individuals to associate freely with others, to practice religion freely, to speak freely, and to possess their persons, homes, and property without fear of arbitrary intrusion.⁸³ Procedural rights include rights that ensure a fair trial and constrain the levying of punishments.⁸⁴

Most abstractly, the Fifth and Fourteenth Amendments incorporate the two great procedural principles of liberty and equality intrinsic to the process of practical reason itself as constitutional constraints on *all* governmental decisionmaking—whether by the legislature, by judges, or by the executive branch—ensuring fundamental procedural fairness in the making, interpretation,

81. *Marbury v. Madison*, 5 U.S. (1 Cranch) 137, 176 (1803); *see also McCulloch*, 17 U.S. at 403-05.

82. *See McCulloch*, 17 U.S. at 403-05 (arguing that the Constitution was promulgated by a convention of delegates elected by state legislatures and submitted for ratification to conventions of delegates “chosen in each state by the people thereof,” and “[f]rom these conventions, the constitution derives its whole authority. The government proceeds directly from the people; is ‘ordained and established,’ in the name of the people; and is declared to be ordained, ‘in order to form a more perfect union, establish justice, insure domestic tranquility, and secure the blessings of liberty to themselves and to their posterity.’ . . . The government of the Union, then . . . is, emphatically and truly, a government of the people. In form, and in substance, it emanates from them. Its powers are granted by them, and are to be exercised directly on them, and for their benefit.”).

83. *See U.S. CONST. amends. I, II, III, IV, V.*

84. *See U.S. CONST. amends. VI, VII, VIII.*

and enforcement of the law.⁸⁵ Indeed, the categorical requirement of practical reason that every person affected by a moral decision must be treated the same as every other similarly situated person may be taken as the core of the concept of equality embodied in the Equal Protection Clause of the Fourteenth Amendment.⁸⁶ And the categorically moral concept of respect for the moral autonomy of each rational person as both maker and object of moral judgments, as well as respect for the intrinsic worth of all persons, may be taken as the essential moral core of the concept of liberty that underlies the Due Process Clause of the Fifth⁸⁷ and Fourteenth Amendments.⁸⁸

But adherence to these two great principles of procedural fairness is not the sole determinant of justice. The people must also be afforded the right to determine the moral objectives of a just society under empirical conditions and the moral means of achieving them within the constitutional constraints they have imposed upon themselves. Thus, the Necessary and Proper Clause of Article One of the Constitution grants Congress the power to promulgate laws to implement the powers delegated to it by the people;⁸⁹ Article Five ensures the people's right to amend the Constitution;⁹⁰ the Ninth and Tenth Amendments secure against intrusion by the State the substantive personal liberties traditionally held by the people but not enumerated in the Constitution;⁹¹ the Tenth Amendment assures to state legislatures the "police power" that protects the public health, safety, welfare, and morals;⁹² and the Fourteenth Amendment

85. See *Bolling v. Sharpe*, 347 U.S. 497, 499-500 (1954) (stating that the concepts of equal protection of laws and due process both stem from the American ideal of fairness).

86. The Fourteenth Amendment provides, in relevant part:

No State shall make or enforce any law which shall abridge the privileges or immunities of citizens of the United States; nor shall any State deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws. . . . The Congress shall have power to enforce, by appropriate legislation, the provisions of this article.

U.S. CONST. amend. XIV, §§ 1, 5; see *Reynolds v. Sims*, 377 U.S. 533, 565 (1964) ("And the concept of equal protection has been traditionally viewed as requiring the uniform treatment of persons standing in the same relation to the governmental action questioned or challenged.").

87. The Fifth Amendment provides, *inter alia*, that "[n]o person shall . . . be deprived of life, liberty, or property, without due process of law." U.S. CONST. amend. V.

88. See *West Coast Hotel Co. v. Parrish*, 300 U.S. 379, 391 (1937) ("[T]he liberty safeguarded [by due process] is liberty in a social organization which requires the protection of law against the evils which menace the health, safety, morals, and welfare of the people.").

89. U.S. CONST. art. I, § 8, cl. 18.

90. U.S. CONST. art. V.

91. The Ninth Amendment provides, "The enumeration in the Constitution, of certain rights, shall not be construed to deny or disparage others retained by the people." U.S. CONST. amend. IX. The Tenth Amendment provides, "The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people." U.S. CONST. amend. X.

92. The Tenth Amendment "police power" is "[t]he inherent and plenary power of a

expressly reserves to Congress the “power to enforce, by appropriate legislation, the provisions of this article.”⁹³ Thus, the process for guaranteeing just government that is fair to all and dedicated to the common good as the people perceive it is built into the Constitution.

The Constitution grants all legislative powers to the legislative branch, empowering Congress to enact general laws to carry out its mandate.⁹⁴ It vests the President and the executive branch with power to execute the laws.⁹⁵ And, finally, it grants the courts jurisdiction to preside over particular cases and controversies⁹⁶ and implies the principle of judicial review through its delegation of authority to judges. In this manner, it effects the intent of the drafters that judges, acting as intermediaries between the people and Congress, review and interpret the laws.⁹⁷

The Constitution thus structures a government of ordered liberties, or positive laws, composed of intrinsically fair structural and enabling principles and concomitant constraints upon the power of the State and each of its offices. Most importantly, it constrains all legislative, executive, and judicial decisionmakers to adhere to the fundamental principles of liberty, equality, and impartiality intrinsic to practical reason and to those principles of ordered liberty deemed by the drafters to be fundamental to a just society and thus written into the Constitution to ensure fairness and justice for all. The Constitution establishes the parameters of the rule of law.

Within the constitutionally structured system of positive laws adopted by the United States, the roles of legislators, who make general laws on behalf of the people, and of judges, who construe and apply the law in the context of particular cases and controversies, are both vital, but they are separate and distinct. Legislators make fair and impartial laws of general application within the constraints of the Constitution. It thus falls to them—in light of their access to broad sources of support and information, their role in an assembly of representatives of all, and the generalized effect of the laws they make—to make sound social policy and to ensure its intrinsic morality by adhering to the requirements of practical reason embodied in the Constitution in its promulgation and enactment. If the law is then interpreted and applied by fair and impartial

sovereign to make all laws necessary and proper to preserve the public security, order, health, morality, and justice. It is a fundamental power . . . subject to due process and other limitations . . . [and can be delegated] to local governments.” BLACK’S LAW DICTIONARY 1196 (8th ed. 2004); *see supra* note 91 (for text of the Tenth Amendment).

93. U.S. CONST. amend. XIV, § 5.

94. U.S. CONST. art. I, § 1.

95. U.S. CONST. art. II, § 1.

96. U.S. CONST. art. III, § 2.

97. *See Marbury v. Madison*, 5 U.S. (1 Cranch) 137, 176-77 (1803); *see also* THE FEDERALIST No. 78, at 467 (Alexander Hamilton) (Clinton Rossiter ed., 1961) (“[T]he courts were designed to be an intermediate body between the people and the legislature in order, among other things, to keep the latter within the limits assigned to their authority. The interpretation of the laws is the proper and peculiar province of the courts.”).

judges using traditional legal reasoning in particular cases and controversies, its intrinsic morality will be preserved, and the laws so produced will be kept moral. For the role of judges in a functioning constitutional democracy is *not* that of policymakers, as some contemporary legal theorists have averred,⁹⁸ but that of interpreters and implementers of constitutional, legislative, and common-law principles and rules within the context of particular cases. And traditionalist judges fulfill that role by reasoning from the facts of the case and the applicable law, in accordance with the principles of practical reason as incorporated into the law, to judgments that instantiate the individual rights and obligations of the parties in a way that is consistent with and becomes part of the positive law to guide future cases in materially similar circumstances. Thus, together with its counterpart (a rational and fair legislative process), traditional judicial reasoning preserves and furthers the ends of practical reason within the constitutional framework.

The system of the positive law in which lawmaking and judicial interpretation take place constitutes a hierarchy of constraints upon the unfettered liberty, both of those who are subject to the law and of those who make and interpret the law. Within this hierarchy, constitutional principles trump statutes, which in turn trump case law, with all being ever subject to change through constitutional amendment, statutory enactment, amendment, and revocation, and judicial interpretation.⁹⁹ And it is within this system that legal reason operates to apply the law under particular circumstances in case after case, adjusting the boundaries of legal concepts on an incremental basis as statutes do on a generalized basis. Case law gives way to statutory law when the need for a uniform general law arises and a statute is promulgated and enacted, but case law is then invoked anew to concretize the application of the statute in particular cases and controversies.¹⁰⁰ The entire body of the law is thus subject to ongoing refinement and modification by concretizing judicial interpretation, evaluation, and judgment. And the opinions judges draft in particular cases enter into a

98. See, e.g., DWORKIN, JUSTICE, *supra* note 3, at 5; see also *supra* note 21 and accompanying text.

99. This system reflects the common-law system inherited from England. See SIR MATTHEW HALE, THE HISTORY OF THE COMMON LAW OF ENGLAND 144-45 (Charles M. Gray ed., 1971) (1713) (describing the decrees of kings as overriding parliamentary statutes, which in turn override the judgments of courts, while private opinion has no legal force whatsoever); see also SIR EDWARD COKE, THE SECOND PART OF THE INSTITUTES OF THE LAWS OF ENGLAND, in COKE ON MAGNA CHARTA [A9-10] (Omni Publ'sns 1974) (1797). Moral principles are embedded in all these sources of the common law. See *id.* at A5, 6, 9-10 (setting out constituents of common law).

100. See THE FEDERALIST No. 78 (Alexander Hamilton), *supra* note 97, at 467 (“A constitution is, in fact, and must be regarded by the judges as, a fundamental law. It therefore belongs to them to ascertain its meaning as well as the meaning of any particular act proceeding from the legislative body. If there should happen to be an irreconcilable variance between the two, that which has the superior obligation and validity ought, of course, to be preferred; or, in other words, the Constitution ought to be preferred to the statute, the intention of the people to the intention of their agents.”).

flexible body of law that grows and changes incrementally with each case, even as it is kept uniform, coherent, reliable and just through the adherence of judges to “strict rules and precedents” and to traditional legal reason in deciding each case.¹⁰¹

The positive law within which judicial decisionmaking functions thus constitutes an autonomous and strictly bounded system of official pronouncements authorized by the Constitution. It is a system that governs the official public life of society and that is shaped by interaction with empirical factual and legal circumstances over time so as to maintain itself and the relation of its parts. It is an autopoietic system.¹⁰² And it is, by adhering to rules and principles in the positive law, within the constraints of practical reason as incorporated into the law and the methodology by which it is applied that judges fulfill their constitutional function of preserving and protecting the boundaries of the laws made by the people (rather than substituting their own private and personal judgments for those sanctioned by due process). Indeed, the integrity and functionality of the system depends upon the shared expectation that lawmakers and judges will play by the rules of the game, i.e., that they will follow the rules and precedents produced by the system itself and will not change the rules to fit their own personal conceptions of the “best” construction of morality or to achieve the “best” social consequences. That is why traditional jurisprudence matters. But how does legal reason, its modus operandi, work in particular cases to achieve these ends?

IV. APPLIED LEGAL REASON: THE LEGAL ANALYTIC

I have argued that the positive law in America is structured and maintained in accordance with moral principles and that its purpose is to further justice and the common good. I also have argued that each legal judgment contributes to an organic empirical system of positive laws that is at once autonomous and strictly bounded and also dynamic, indeterminate, and open-ended and that the formal process by which traditional legal judgments are made is not only objective and rational in form, but also particularized, case-bound, interpretive, evaluative, and prescriptive in application and, therefore, concretized. And I have argued that the role of judges is to interpret and apply the law in particular cases fairly and rationally in accordance with its intent so as to do justice to the parties under the law and to preserve the rule of law. But I have not yet indicated how traditional legal reason works within the context of a particular case so that the judgment reached may be said to be best to effect justice in that case and to maintain the purpose, integrity, and functionality of the law overall—and, therefore, to be one that should be implemented. I have not yet addressed the *legal analytic*, or model of traditional judicial decisionmaking within the confines of a particular case.

101. *Id.* at 471.

102. See *supra* note 77 and accompanying text.

*A. Traditional Judicial Methodology*¹⁰³

A traditional judicial opinion customarily begins with a brief introduction stating the nature of the case, the stage of the proceedings, and a very brief statement of the legal issues within the positive law that are to be decided; it may also announce the disposition. The introduction is followed by a statement of facts necessary to the disposition, whether procedural or substantive. These sections set the legal and factual scope of the opinion.

The opinion then sets out the standard of review. These standards—such as abuse of discretion; legal or factual sufficiency of the evidence; de novo review of legal issues; or, in a case challenging the constitutionality of a law, the compelling state interest or rational basis test—establish norms for evaluating the trial court's judgment. Is the case ripe for adjudication? Is the defendant state actor immune from suit? Is deference to be paid to the trial judge's decision so long as it is rational? Is there a presumption in the applicable law in favor of one party or the other, or in favor of innocence over guilt, or in favor of a conviction for a lesser offense over a greater offense in case of reasonable doubt? Is the presumption rebuttable? Does affirmance require proof of a factual scenario by a preponderance of the evidence, or is the judgment reversible only if it is contrary to the overwhelming weight of the evidence or if it is supported by no more than a scintilla of evidence? These are all either objective or rational tests that are designed to determine pure questions of law, such as questions of duty or jurisdiction, or semi-subjective evidentiary or balancing tests that instruct the court how to weigh the evidence in the case to determine whether it is sufficient to support the judgment.

The next section of the opinion, the discussion, sets out each legal issue and the substantive or procedural law that governs each and analyzes the case accordingly. The statement of the law provides premises in the form of legal rules, rules of construction, and standards from and by which courts can reason logically from the legal issues posed by a case to an opinion and judgment—whether the issue regards the interpretation of a constitutional principle, statutory provision, contract, or will, or the proof of the elements of a crime, claim, or affirmative defense. The rest of the discussion consists of the application of the law to the facts of the particular case under the applicable rules and standards as stated in statutes, rules, and precedents so that a conclusion (or holding) may be reached that rationally and fairly decrees the parties' rights and obligations in a manner consistent with prior law and a judgment may issue that prescribes the actions to be taken to implement the holding.

This description of judicial decisionmaking—or applied judicial reasoning—seems incontrovertible and uncontroversial. It describes what judges actually do when they are reasoning legally. The devil, however, is in “the application of the law to the facts,” or in the interpretation of the governing rules of law and the evaluation of alternatives so that the “best” judgment is made, or

103. This section analyzes intermediate appellate opinions, but the analysis is generally applicable to all judicial opinions.

that which is most coherent with the body and purpose of the law and most conducive to the good of the whole, consistent with the principles of fairness and justice. This is where the anti-traditionalist argument comes into play. For if traditional jurisprudence derives the answers to legal questions by mere deduction from backward-looking factual reports, as the anti-traditionalist argument says it does, then, by definition, traditional jurisprudence has no principled methodology for deciding among alternative interpretations of a legal concept, or for deciding among possible alternative outcomes, or for reaching forward-looking conclusions in a changing social landscape. According to the anti-traditionalists, traditional jurisprudence cannot resolve “hard cases.” Similarly, if traditional jurisprudence is ruleless, where is the bar to introducing rules, or theory-driven principles, to bring principled structure to what is otherwise mere casuistry? Not surprisingly, therefore, legal philosophers seek to justify alternative strategies to traditional legal reasoning as producing *better* law than traditional legal reasoning within the constraints of the positive law, and even as producing *some* law where they contend the positive law, being fixed and backward-looking, has gaps.

It is here that Dworkin advocates that judges—especially Supreme Court Justices—should implement the objectively true moral principles that underlie the positive law.¹⁰⁴ It is also here that Posner advocates that judges of integrity should apply a “pragmatic” utilitarian analysis¹⁰⁵ and that Scalia argues that judges should implement the “original meaning of the text.”¹⁰⁶ And it is here that Sunstein suggests judges may adhere to whatever strategy they choose, so long as they decide most cases on lower- or mid-level principles on which they agree, introducing fundamental principles only in the most important cases, and thus minimizing the harm done to the fabric of the law by divisive judgments derived from theory.¹⁰⁷ Thus, in the next section, I address the process by which traditional judges actually do apply the law to the facts to show that there are no “gaps” in the positive law they apply and no “hard” cases that traditional legal reasoning fails to resolve, justifying resort to alternative strategies of judicial decisionmaking.

B. Traditional Judicial Analysis

The essential difference between a traditional jurist and a theory-driven jurist is that rather than taking an active constructive approach to the law, a traditional judge decides cases on the understanding that the role of a judge is to further the purpose of the laws that the people themselves have made through their representatives and ratified by their consent or that have evolved through the incremental process of case law. This approach requires judges of integrity to interpret and apply the law within existing constraints in the positive law and in

104. See, e.g., DWORKIN, FREEDOM’S LAW, *supra* note 17, at 73-74, 343.

105. See POSNER, PROBLEMATICS, *supra* note 6, at xi-xii.

106. See SCALIA, *supra* note 9, at 45.

107. See SUNSTEIN, LEGAL REASONING, *supra* note 26, at 5.

light of the particular circumstances of the case in such a way as to maintain the integrity and functionality of the law and to effect the common good as the people have defined it. So what are the constraints the positive law imposes upon judicial interpretation?

At the most general and abstract level of the positive law in the United States are constitutional and traditionally held fundamental common-law principles deemed essential to the concept of ordered liberty itself. These abstract principles are the great enabling and constraining concepts that shape the contours of the positive law, but their empirical scope is almost entirely undefined in the Constitution itself.¹⁰⁸ Subordinate to these are rules of law that define the limits, or extensions, of factual situations that fall within the scope of general principles. Rules thus define the relationship between broad abstract intellectual concepts and their practical application to the facts in different sets of circumstances.

In all “hard cases,” or cases requiring the exercise of judicial discretion, the judge’s first task is to determine the scope of unclear or overlapping legal concepts, where “the extension of the word or phrase is all and only those things that satisfy the descriptions speakers associate with the term(s),”¹⁰⁹ in light of the totality of the material circumstances of the particular case or controversy and the relevant principles, rules, and precedents in the positive law. Then the judge’s task is to derive the logical conclusions from that interpretation under the appropriate standards and to evaluate alternatives to find the best—or most logical and functional—fit between the law applied, the purpose for which it was intended, and the consequences for the parties and for the law itself.

Contrary to the portrayal of traditional or “conventional” legal reason by rationalist legal theorists as being “mere deduction” from conventionally fixed principles, the extension of legal concepts (or principles and rules of law) is not fixed, and the best judgment cannot be determined solely by deductive reasoning. Nor are judges licensed by the Constitution to interpret legal concepts constructively in accordance with some “justification beyond law’s warrant, beyond any requirement of consistency with decisions made in the past,” finding justification for their decisions in “abstract justice, or in the general interest, or in some other forward-looking justification.”¹¹⁰ Rather, in traditional legal

108. Chief Justice Marshall recognized this essential feature of the American Constitution in *McCulloch v. Maryland*, 17 U.S. (4 Wheat.) 316, 407 (1819) (“A constitution, to contain an accurate detail of all the subdivisions of which its great powers will admit, and of all the means by which they may be carried into execution, would partake of the prolixity of a legal code, and could scarcely be embraced by the human mind. . . . Its nature, therefore, requires, that only its great outlines should be marked, its important objects designated, and the minor ingredients which compose those objects, be deduced from the nature of the objects themselves.”).

109. See David O. Brink, *Legal Interpretation, Objectivity, and Morality*, in OBJECTIVITY, *supra* note 18, at 12, 21. Brink notes that “[t]he claim that general [legal] terms are open textured . . . fits with[in] . . . a semantic tradition that includes [philosophers from] John Locke . . . [to] Rudolph Carnap.” *Id.* at 21.

110. DWORKIN, LAW’S EMPIRE, *supra* note 2, at 115.

reasoning, the extension of legal concepts is generally determined by what can broadly be termed associative and dissociative reasoning within the constraints of the positive law. Once this is determined, principles may be reconciled and interests evaluated to determine the best outcome for the parties and the law within the confines of the legal process itself.

In reasoning legally, the traditionalist judge determines what kinds of situations fall within the scope of each concept by looking first to past interpretations in materially similar and dissimilar situations—precedents—and analogizing to the material elements of those cases or distinguishing them on the same basis. The judge uses these tools to draw or reshape the boundaries of each concept so as best to accommodate the material facts of the case, moving outward from the particular to the more general as necessary to accommodate new data and to find the least general rule that will accommodate the facts while maintaining the structural integrity of the law at issue.¹¹¹ Often, this process suffices to determine that the case falls squarely within the scope of a particular rule. When it does not, the judge has recourse to judicially and legislatively created tools of construction, including the plain language of the text to be interpreted, the harmonization of all its parts, the intent of the drafters, and, where appropriate, the policy of the state.¹¹² When the best interpretation of the governing rules is determined, the judge reasons deductively from that interpretation, under appropriate evidentiary or legal standards, to whatever judgment logically follows. The objective of this process is to ensure that the

111. See CHARLES FRIED, SAYING WHAT THE LAW IS: THE CONSTITUTION IN THE SUPREME COURT 24-25 (2004) (“It is the way of doctrine to develop by analogy, moving out from a core of concrete instances by progressive, metaphorical extensions until a new and more abstract statement is achieved.”). Lloyd Weinreb has likewise analyzed the role of analogy in legal reason. See generally LLOYD L. WEINREB, LEGAL REASON: THE USE OF ANALOGY IN LEGAL ARGUMENT (2005). On Weinreb’s account, when a judge is uncertain about how to legally classify some phenomenon, he reasons by analogy, or employs “abductive” reasoning, which makes a conclusion possible without making it certain, unlike deductive reasoning. *See id.* at 20-24. Weinreb supports his claim that legal reasoning encompasses analogical reasoning as well as deduction with the argument (with which I agree) that “[u]nless one is able to identify an object as a member of a class despite its differences from other members of the class, no deductive inference is possible.” *Id.* at 127. Sunstein likewise has argued that analogical reasoning plays a central role in legal reasoning, according it a much less constrained role than Weinreb or Fried. See SUNSTEIN, LEGAL REASONING, *supra* note 26, at 62-67.

112. For example, the Texas Government Code instructs that judges construing statutes may consult, *inter alia*, the: “(1) object sought to be obtained; (2) circumstances under which the statute was enacted; (3) legislative history; (4) common law or former statutory provisions, including laws on the same or similar subjects; [and] (5) consequences of a particular construction.” TEX. GOV’T CODE ANN. § 311.023 (West, Westlaw through 2009 Reg. Leg.). It further instructs judges to presume: “(1) compliance with the constitutions of . . . [the] state and United States is intended; (2) the entire statute is intended to be effective; (3) a just and reasonable result is intended; . . . and (5) public interest is favored over any private interest.” *Id.* § 311.021. The Code also advises judges that the rules it contains “are not exclusive.” *Id.* § 311.003.

judgment made is at once fair to the parties and consistent with fairness. Furthermore, it ensures that the judgment is most conducive to the integrity, functionality, and purpose of the positive law in which it takes its place, and thus most conducive to furthering the common good.

If alternative interpretations of a term, rule, or principle material to the case are rationally justifiable, such that different prospective judgments follow logically from different interpretations, the consequences of those prospective judgments for the parties and the law are assessed by the judge to maintain the best fit with the function and purpose of the law. The judge determines the foreseeable effects that the potential decision may have upon the parties and upon the functionality and vitality of the body of the law by weighing alternatives under applicable constitutional constraints, rules, standards, and relevant legislative policies. The “best” interpretation and application of a legal concept is that which most closely reflects the ordering of the value-laden concepts built into the law as it has been defined, refined, and redefined over time within the law itself to maintain the functionality and integrity of the law and effect its purpose fairly and rationally. The judge need not move beyond the constraints of the law to decide the case in accordance with the dictates of moral theory or social utility as he “best” construes them. Nor does the judge possess any such warrant under the law.

For example, the principle that there must be both offer and acceptance to create a binding contract is a fundamental principle of contract law. By itself, the principle is abstract, universal, and objective, and it sets the contours of contract law—but it is not in itself linked to any particular set or sets of factual circumstances. Rather, what constitutes offer and acceptance in different sets of circumstances and what does not is determined in particular case after particular case, and the law of contract is developed through repeated acts of interpretation of precedents, rules, statutes, logical arguments, and the weighing of alternative outcomes under applicable standards. The repeated interpretation and application of principles of law, such as offer and acceptance, in different sets of legal and factual circumstances establishes and readjusts the boundaries and weights of legal rules vis-a-vis each other. And the same process holds even if what is at issue is not the formation of a contract or the effectuation of a tort, but the construction of a state or federal statute.¹¹³

The interpretive and evaluative constraints of the positive law upon adjudication apply even at the level of constitutional construction upon which most jurisprudential theory concentrates. For example, in construing the Equal Protection Clause of the Constitution,¹¹⁴ no court cognizant of its own responsibility to maintain the functionality, integrity, and purpose of the law—and no judge cognizant of his oath to uphold the laws of the United States or of his accountability to the parties and the law—could simply ignore the language of the clause, the precedents constraining its interpretation, or the facts of the case before him. Similarly, no judge of integrity ignores the traditional

113. See 29 U.S.C. §§ 1001-1461 (2006 & Supp. 2009).

114. U.S. CONST. amend. XIV, § 1.

comparative and deductive tools of legal interpretation, evaluation, and prescription—such as those rules of construction condemning intentionally discriminatory, overbroad, or constitutionally vague statutes, or requiring a rational relationship between the language of the statute and a constitutionally legitimate state interest, or requiring that a statute be “narrowly tailored” to further a “compelling” state interest. No court or judge cognizant of its responsibility acts independently of these constraints.

The repeated exercise of legal reason by judges in particular cases and controversies defines and redefines the boundaries of the liberties the members of a group or society are free to exercise against each other and the constraints of the positive law upon those liberties. And together with the texts of the Constitution, statutes, and precedents, this case-by-case analysis creates a dynamic body of laws that function organically together as an ongoing, self-creating, and self-adjusting social compact—just as the process of making repeated moral judgments in different circumstances within a body of governing moral principles establishes and refines the content of moral precepts and demarcates the limits of their application, generating a coherent and functional moral system over time. Needless to say, however, any legal judgment that depends upon the interpretation of legal concepts or the evaluation of outcomes under empirical standards, rules, and principles in different circumstances over time will be to some degree subjective and indeterminate. Thus, although flexible, it will be subject to correction by successor or higher courts, or to modification or overruling by legislatures, so that the law remains simultaneously bounded by systemic constraints and capable of growth.

Rules of law, like moral rules, are thus *not* fixed and backward-looking, but dynamic, open-ended, and indeterminate. And the opinions and judgments that define and refine them are not “ruleless,” as Sunstein contends,¹¹⁵ or “conventional” and “backward-looking factual reports,” as Dworkin contends,¹¹⁶ simply because the rules are not fixed and determinate. Nor is traditional jurisprudence incapable of resolving “hard cases” because it is not theory-bound. Rather, when the creation of the positive law through the exercise of legal reason in particular cases is understood correctly as an ongoing rational and evaluative empirical process of making concrete legal judgments and incorporating them into the dynamic body of the positive law, it is evident that “hard cases,” like moral dilemmas, are simply those cases that present an empirical nexus in which the scope of two or more legal concepts overlap or the scope of one or more legal concepts is unclear under the circumstances of the case, requiring resolution by interpretation and evaluation within the context of the positive law. In short, *all* cases that present genuine material issues of law or fact are “hard cases.”

Contrary to contemporary jurisprudential dogma, legal reason as traditionally employed by judges in the Anglo-American constitutional, statutory, and common-law tradition of adjudicating cases and controversies is a complex form of practical reason employed on an ongoing basis within the confines of the

115. See SUNSTEIN, *LEGAL REASONING*, *supra* note 26, at 5.

116. See DWORKIN, *LAW'S EMPIRE*, *supra* note 2, at 225.

positive law to which each judgment contributes and which each modifies. When its actual nature is recognized, it is clear that there are no “hard cases” that evade resolution by traditional legal reasoning or require resolution by recourse to constructive sources of “true” or “better” law. There are no “gaps” that judges must actively seek to fill by recourse to theory from outside the positive law. There is therefore no justification for judges to adopt alternative strategies to traditional jurisprudence to resolve “hard cases.”

V. ALTERNATIVE JUDICIAL STRATEGIES

No scholar, to my knowledge, has given an argument for when and why the exercise of traditional legal reason ceases to provide a satisfactory methodology for resolving cases so that “strategic” alternatives to traditional jurisprudence are justified, other than the anti-traditionalist argument, which fails. And no contemporary legal philosopher, to my knowledge, has produced an alternative strategy for resolving legal cases that is at once comprehensive, dynamic, and capable of producing concretizing judgments that will maintain the purpose, integrity, and functionality of the rule of law over time, as traditional jurisprudence does. Rather, the results of the application of alternative strategies demonstrate their inherent tendency to undermine the positive law and the social compact it implements.

Sunstein, for example, accepts the rationalist definition of “rules” as fixed and determinate “approaches to law that aspire to make legal judgments in advance of actual cases.”¹¹⁷ He then construes traditional jurisprudence—in which rules are *not* fixed and determinate—as “ruleless.”¹¹⁸ Thus, while he recognizes that “[a]ll rules are defined in terms of classes,”¹¹⁹ he does not explore this concept as it applies to traditional legal interpretation. More importantly, he posits a continuum between “ruleless” traditional jurisprudence and “rule-bound justice,” or theory-driven jurisprudence, which he defines as holding that judges “should avoid open-ended standards or close attention to individual circumstances” and “should attempt instead to give guidance to lower courts, future legislators, and ordinary citizens through clear, abstract rules laid down in advance of actual applications.”¹²⁰ This continuum then becomes the linchpin of his argument for minimalism, the doctrine that judges whose decisions are driven by disparate and intractable theories can—and, as principled judges, even *should*—converge on lower-level or mid-level principles when they have profound disagreements of principle, but should resort to rule-bound justice to

117. SUNSTEIN, *LEGAL REASONING*, *supra* note 26, at 21 (emphasis omitted).

118. *See id.* at 13 (“Much of what lawyers know is a set of practices, conventions, and outcomes that is hard to reduce to rules, that sometimes operates without being so reduced, and that is often just taken for granted. This background knowledge makes legal interpretation possible, and it sharply constrains legal judgment.”).

119. *Id.* at 24.

120. *Id.* at 10 (emphasis omitted).

decide the most important cases.¹²¹

The problem with Sunstein's theory is that the important cases are decided by principled judges in advance of actual cases and on general, and admittedly divisive, grounds—the exact opposite of the constitutional mandate that judges must decide particular cases or controversies in such a way as to maintain the integrity and functionality of the law.¹²² Ultimately, therefore, minimalist legal theory collapses into advocacy of a system of government by judicial decree in which the ideological might of a bare majority of the judges on a court of last resort makes right.

Similarly, Dworkin's perfectionist conception of the law as legitimized by its logical derivation from objectively true moral propositions leads to an entirely different set of legal judgments from those produced by traditional legal reasoning—namely, political judgments as opposed to traditional legal judgments.¹²³ Unlike traditional jurisprudence, perfectionist jurisprudence does not begin with the concept that the positive law is a dynamic set of indeterminate empirical principles that is intrinsically moral and is kept moral through a fair and rational process of legal reasoning to which each legal judgment contributes. It begins with the axioms that the fundamental principles of democracy are objectively true moral propositions¹²⁴ independent of the positive law and that the “best” construction of the extension of constitutional concepts is given directly by moral and political philosophy.¹²⁵ Thus, a judge who, in deciding particular cases, construes and implements the principles of liberty and equality in the Constitution in accordance with his own “best” construction of the rational requirements of the abstract principles of liberty and equality participates in constructing the “true” conditions of a just democratic society. What perfectionists present as an objectively true construction of constitutional principles, however, necessarily reduces to a profoundly *subjectivist* view of legal “truth.” For the values in the positive law are exchanged for the values the

121. Sunstein argues that judges decide most cases on such “lower-level principles,” reasoning largely by analogical reasoning, which does not require anything like “horizontal and vertical consistency,” but only “[l]ocal consistency,” where there is no pre-existing rule and by interpretive practices within the legal community that frequently take the form of “background knowledge,” although “no particular approach” is required. *See id.* at 32-33, 52, 65.

122. *See U. S. CONST. art. III, § 2, cl. 1* (limiting judicial power to cases or controversies).

123. *See DWORKIN, JUSTICE, supra* note 3, at 2 and accompanying text (advocating perfectionist jurisprudence and stating that it is “particularly important” that philosophers and judges determine what conditions must hold so that legal judgments may follow deductively from objectively true propositions of law “in political communities like our own in which important political decisions are made by judges who are thought to have a responsibility to decide only as required or licensed by true propositions of law”). I develop the argument that perfectionist jurisprudence leads to different judgments from traditional jurisprudence in another article. *See Keyes, Two Conceptions of Judicial Integrity, supra* note 62, at 286-90.

124. *See, e.g., DWORKIN, FREEDOM'S LAW, supra* note 17, at 30-32; *see also supra* note 20 and accompanying text.

125. *See DWORKIN, FREEDOM'S LAW, supra* note 17, at 34-35.

theory puts forward, and the truth value of legal propositions depends entirely upon the “best” contemporary construction of those principles as understood by judges freed of the constraints present in the positive law. Thus, perfectionism ultimately reduces to the personal and private political notions of fallible—and possibly even malleable—individual judges, as even Dworkin, the leading proponent of rational perfectionism, acknowledges.¹²⁶ Accountability is sacrificed to unfettered judicial independence.

Process-oriented judicial decisionmaking in accordance with pragmatist strategies, such as those of Posner or Eskridge, by contrast, need not have consequences for the law as severe as those of perfectionist or other objectivist theories. Taking consequences into account and making morally optimal empirical decisions—those that contribute most to the flourishing life of society—are here seen as integral parts of the process of legal reasoning. But when consequentialism becomes unmoored from the constraints of the positive law and turns to social constructivism based on facts, theories, or principles foreign to the facts of the case and the positive law of the society, it departs from traditional jurisprudence and its virtues. For example, when a court justifies a legal decision by reference to foreign law in a domestic case or by reference to “trends” in “developing” subordinate laws; or when it justifies a criminal sentence on economic grounds or a civil decision on policy grounds, independent precedent, or legislative intent; or when it justifies a legal decision on the basis of judge’s assessment of the “social utility” of the judgment, as Posner advocates;¹²⁷ or even when it resorts to Eskridge’s “critical pragmatism”,¹²⁸ and permits “an ongoing reevaluation of practice, especially dividing practices that marginalize groups of citizens and interest group distributions at the expense of the public interest” in accordance with the judge’s own values, the principles of legal reason have been abandoned and the “pragmatism” that results has all the damaging consequences for the moral, legal, and political social structure as any other theory-driven jurisprudence.¹²⁹ Thus, pragmatism, like perfectionism, fails to offer a sound strategic alternative to traditional jurisprudence.

Indeed, of all the theories of jurisprudence on the current scene, the only theory that defines sound legal strategy in a way acceptable to traditional judges is classical liberalism—not quite as Eskridge describes it,¹³⁰ but as it truly exists. Classical liberal jurists do not, as Eskridge states, “view[] government as a social contract among autonomous individuals who in the distant hypothetical past gave up some of their freedom to escape the difficulties inherent in the state of nature,” legitimizing “laws enacted by prescribed constitutional procedures . . .

126. See *id.* at 74. Dworkin argues that the Constitution makes Supreme Court Justices the moral arbiters of society, accountable only to their own independent “best” construction of the truth, even as he acknowledges their fallibility.

127. POSNER, PROBLEMATICS, *supra* note 6, at xi.

128. ESKRIDGE, *supra* note 42, at 193 (emphasis omitted).

129. As noted above, Eskridge himself recognizes this danger. See *supra* notes 58-60 and accompanying text.

130. See *supra* notes 47-48 and accompanying text.

by the consent expressed in the original social contract.”¹³¹ Rather, classical liberals—among whom I count myself—view constitutional democratic government as an intrinsically moral dynamic social compact of ordered liberties, or positive laws. The social compact is grounded in a foundational Constitution that incorporates fundamental principles of liberty, equality, and impartiality; separates delegated governmental powers among executive, legislative, and judicial branches; and charges judges with interpreting the law in particular cases and controversies within the constraints of the rule of law. The people consent to the rule of law thus established by their ratification of its prescriptions and their participation in the political process.¹³² They do not look back to the eighteenth century to determine the fixed meaning of constitutional terms, as originalists do.¹³³ Important policy issues are, indeed, properly resolved by “the legislature or some other majoritarian institution” on this theory,¹³⁴ as Eskridge states. But legal judgments are constrained by the entire hierarchical body of the positive law as it is defined and redefined over time, by the facts of the particular cases in which the law is applied, and by the principles of practical reason that constrain both legal reason and the positive law. Thus, the jurisprudence by which the social compact is maintained is, like the compact itself, both dynamic and concretizing and intrinsically just. Indeed, an integral part of the judge’s role in classical liberal theory is to ensure that policies made by majoritarian legislatures are interpreted and applied in accordance with the rule of law, and not in accordance with the unfettered will of temporary legislative majorities unbound from the constraints of ordered liberty or in accordance with the independent will of judges. Traditional jurisprudence is classical liberal jurisprudence rightly understood.

CONCLUSION

I have argued that contemporary legal philosophers generally misconstrue the nature of traditional jurisprudence. And I have argued that they and their adherents rely on that misapprehension to justify the abandonment of sound judicial decisionmaking that maintains the justice, integrity, and functionality of the rule of law in favor of alternative judicial strategies that actually undermine the strength of the rule of law. The positive law of the United States is not a set of backward-looking factual reports or mere conventions, as Dworkin and others would have it. Nor is it a set of lower-level and mid-level principles at the low end of a continuum between mere “casuistry” and principled, theory-driven jurisprudence, as Sunstein opines. Likewise, traditional legal reasoning is neither

131. ESKRIDGE, *supra* note 42, at 111.

132. I develop the concept of contemporary classical liberalism in another article. See generally Keyes, *The Just Society and the Liberal State*, *supra* note 62.

133. See *id.*; see also SCALIA, *supra* note 9, at 45 (“[T]he originalist at least knows what he is looking for: the original meaning of the text. Often—indeed, I dare say usually—that is easy to discern and simple to apply.”).

134. See ESKRIDGE, *supra* note 42, at 133.

mere logical deduction from fixed conventional legal principles that cannot supply answers to hard questions nor “rulelessness” that principled judges must move beyond in important cases of constitutional principle. Nor is traditional jurisprudence merely one strategy among many for resolving legal cases. Rather, as traditional jurisprudence recognizes, the positive law in a constitutional republic founded, like our own, on principles of liberty and equality for all under laws made by all and for all constitutes a dynamic social compact grounded in intrinsically moral principles of ordered liberty. And legal reason, as traditionally understood and employed by traditionalist judges, is a form of practical or applied moral reason that operates over time within the constitutional framework of the positive law fairly and rationally, concretizing legal principles in particular judgments that further the beneficent purpose of the law, thereby ensuring justice to the parties and contributing to the maintenance of the rule of law.

By adopting the anti-traditionalist argument and encouraging judges to reach outside the positive law and traditional legal reasoning and to implement constructivist conceptions of “better” law,¹³⁵ rationalist legal theorists justify legal decisionmaking that violates the requirements of practical reason, flouts constitutional safeguards, and destabilizes the law. Traditional jurisprudence—or classical liberal jurisprudence—alone accommodates the dynamism of the law and society and concretizes the law in particular cases while maintaining the purpose, integrity and functionality of the rule of law over time. It thus commends itself as the model of choice for ensuring justice through valid and sound legal decisionmaking in all cases and controversies.

135. As discussed previously in this paper, I use the term “better” law to mean that which is derived from moral or political theory, or “pragmatically” directed to the maximization of social utility or to the engineering of social ends.

A CASE OF DOUBTFUL CERTAINTY: THE COURT RELAPSES INTO SEARCH INCIDENT TO ARREST CONFUSION IN *ARIZONA V. GANT*

GEORGE M. DERY III*

INTRODUCTION

The Supreme Court has long recognized the need to craft clear rules to guide police in their daily work because the Fourth Amendment cannot control officers who do not understand it.¹ In *New York v. Belton*, a case in which the Court enabled police to search as incident to arrest the passenger compartment of an arrestee's vehicle,² the Court expressed wariness of a "highly sophisticated set of rules, qualified by all sorts of ifs, ands, and buts and requiring the drawing of subtle nuances and hairline distinctions" because such laws "may be 'literally impossible of application by the officer in the field.'"³ *Belton*'s practical concern in forming a workable rule led the Court to a generalized rule that all passenger compartments fell within search incident to arrest.⁴

Despite such an effort at clarity, search incident to arrest—at least in the vehicle context—became a doctrine divorced from its reasonable moorings. Nearly a quarter century after *Belton*, in *Thornton v. United States*, the Court upheld a search incident to arrest of a vehicle even though its driver had been "handcuffed and secured in the back of a squad car" at the time of the search.⁵ *Thornton*'s arrestee, fettered and in police custody, hardly appeared to be in a position to retrieve a weapon or evidence, the concern originally addressed by the Court's search incident to arrest precedent.⁶ *Thornton* mystified Justice Scalia, who believed the Court's holding envisioned an arrestee "possessed of the skill of Houdini and the strength of Hercules."⁷

Four years after *Thornton* strained search incident to arrest to near breaking, the Court heard *Arizona v. Gant*.⁸ The *Gant* Court, wishing to return to "the

* Professor, California State University Fullerton, Division of Politics, Administration, and Justice. Former Deputy District Attorney, Los Angeles, California; J.D., 1987, Loyola Law School, Los Angeles, California; B.A., 1983, University of California, Los Angeles.

1. See *New York v. Belton*, 453 U.S. 454, 458 (1981). The Fourth Amendment provides: The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

U.S. CONST. amend. IV.

2. *Belton*, 453 U.S. at 462-63.

3. *Id.* at 458 (citation omitted).

4. *Id.* at 460.

5. *Thornton v. United States*, 541 U.S. 615, 625 (2004) (Scalia, J., concurring).

6. See *Chimel v. California*, 395 U.S. 752, 763 (1969).

7. *Thornton*, 541 U.S. at 626 (Scalia, J., concurring) (quoting *United States v. Frick*, 490 F.2d 666, 673 (5th Cir. 1973)).

8. 129 S. Ct. 1710 (2009).

safety and evidentiary justifications” underlying search incident to arrest’s scope, held that “*Belton* does not authorize a vehicle search incident to a recent occupant’s arrest after the arrestee has been secured and cannot access the interior of the vehicle.”⁹ *Gant* instead offered a new rule for search incident to arrest of vehicles, holding that police could “search a vehicle incident to a recent occupant’s arrest only when the arrestee is unsecured and within reaching distance of the passenger compartment at the time of the search.”¹⁰ To obtain the five votes needed to support this rule, however, the Court apparently had to do some horse trading. To persuade Justice Scalia to join the majority, the Court added a second and entirely novel police right to search incident to arrest: “we also conclude that circumstances unique to the vehicle context justify a search incident to a lawful arrest when it is ‘reasonable to believe evidence relevant to the crime of arrest might be found in the vehicle.’”¹¹ Thus, in an attempt to meaningfully limit search incident to arrest in the vehicle context, *Gant* created a new two-part rule.¹²

The *Gant* rule, however, may lead to further confusion in this troubled area of Fourth Amendment litigation. After a review of the historical background of search incident to arrest in Part I and a consideration of *Gant*’s facts and the Court’s ruling in Part II, this Article critically examines the concerns raised by the Court’s new rule. Those problems could be significant. By failing to adequately define its “reaching distance” limit,¹³ *Gant* could create misunderstandings of Fourth Amendment protections, causing officers to improperly execute its new standard. Further, *Gant*’s allowance of a search upon a “reason to believe” a vehicle contains “offense-related evidence”¹⁴ could expand search incident to arrest beyond its original justifications to become coterminous with the automobile exception. Finally, by failing to specify the “unique” circumstances the Court found in the automobile context,¹⁵ *Gant* might enable future cases to spread its contagion of confusion to search incident to arrest cases outside the vehicle setting.

9. *Id.* at 1714.

10. *Id.* at 1719.

11. *Id.* (quoting *Thornton*, 541 U.S. at 626) (Scalia, J., concurring)).

12. *Id.* at 1726 (Alito, J., dissenting) (“a police officer who arrests a vehicle occupant or recent occupant may search the passenger compartment if (1) the arrestee is within reaching distance of the vehicle at the time of the search or (2) the officer has reason to believe that the vehicle contains evidence of the offense”).

13. *Id.* at 1721 (majority opinion).

14. *Id.* at 1719.

15. *Id.* at 1714.

I. HISTORICAL BACKGROUND OF SEARCH INCIDENT TO ARREST

A. A Hint Became a Suggestion, Was Loosely Turned into Dictum, and Was Finally Elevated to a Decision

The search incident to arrest exception to the Court's warrant requirement has earned a secure position as a solid law enforcement right in Court precedent.¹⁶ The Court has exalted the police right to search incident to arrest as "always recognized under English and American law"¹⁷ and as a rule "not to be doubted."¹⁸ Such an unquestioned authority, however, has a curiously dubious origin.¹⁹ Justice Frankfurter once criticized a particular application of search incident to arrest as proof of "how a hint becomes a suggestion, is loosely turned into dictum and finally elevated to a decision."²⁰ The whole of the search incident to arrest rule itself, with its "sketchy" history,²¹ has followed a similar trajectory.²² Indeed, one commentator has charitably described search incident to arrest's "historical provenance" as "not so clear."²³

Justice Cardozo traced the search right "back beyond doubt to the days of the hue and cry, when there was short shrift for the thief who was caught 'with the mainour,' still 'in seisin of his crime.'"²⁴ This theory placed the creation of search incident to arrest "in early Anglo-Saxon law,"²⁵ making it part of "a loud outcry with which felons . . . were anciently pursued, and which all who heard it were bound to take up, and join the pursuit, until the malefactor was taken."²⁶ Thus, search incident to arrest presumably began as a right of angry villagers,

16. See *United States v. Robinson*, 414 U.S. 218, 226 (1973).

17. *Weeks v. United States*, 232 U.S. 383, 392 (1914), overruled by *Mapp v. Ohio*, 367 U.S. 643 (1961).

18. *Agnello v. United States*, 269 U.S. 20, 30 (1925).

19. For an illuminating and cogent discussion of the history of search incident to arrest, see James J. Tomkovicz, *Divining and Designing the Future of the Search Incident to Arrest Doctrine: Avoiding Instability, Irrationality, and Infidelity*, 2007 U. ILL. L. REV. 1417, 1421-45.

20. *United States v. Rabinowitz*, 339 U.S. 56, 75 (1950) (Frankfurter, J., dissenting), overruled in part by *Chimel v. California*, 395 U.S. 752 (1969).

21. *Robinson*, 414 U.S. at 232.

22. The Court has noted, "As Mr. Justice Frankfurter commented in dissent . . . the 'hint' contained in *Weeks* was, without persuasive justification, 'loosely turned into dictum and finally elevated to a decision.'" *Chimel*, 395 U.S. at 760 (citation omitted).

23. Wayne A. Logan, *An Exception Swallows a Rule: Police Authority to Search Incident to Arrest*, 19 YALE L. & POL'Y REV. 381, 385 (2001). This article provides an interesting overview of the early history of search incident to arrest. See *id.* at 385-90.

24. *People v. Chiagles*, 142 N.E. 583, 584 (1923) (citing 2 POLLOCK & MAITLAND, HISTORY OF ENGLISH LAW 577, 578 (1927)). "[T]he thief caught [with stolen goods in his possession is said to be taken] with the mainour[,] meaning the property *in manu*—‘in his hands.’ BLACK’S LAW DICTIONARY 859 (5th ed. 1979).

25. *Rabinowitz*, 339 U.S. at 72 (Frankfurter, J., dissenting).

26. BLACK’S LAW DICTIONARY, *supra* note 24, at 667.

who, upon capturing a fleeing thief, were empowered to recover stolen property from his person.

Search incident to arrest was quite different by the time of the Framers. Since this search right was asserted by government officials rather than villagers, it was constrained within the “limited powers enjoyed by eighteenth century authorities to execute arrests, especially without a warrant.”²⁷ The Framers were reluctant to “extend discretionary authority” to officials at the scene²⁸ because magistrates were viewed as “more capable than ordinary officers of making sound decisions as to whether a search was justified.”²⁹ This, interestingly, might have been due in part to class bias.³⁰ Framing era commentators “sometimes expressed outright disdain for the character and judgment of ordinary officers,”³¹ believing

[i]t was disagreeable enough for an elite or middle-class householder to have to open his house to a search in response to a command from a high status magistrate acting under a judicial commission; it was a gross insult to the householder’s status as a “freeman” to be bossed about by an ordinary officer who was likely drawn from an inferior class.³²

Therefore, officials performing warrantless arrests for even the most serious offenses—felonies—were protected from civil suit only if the arrestee was actually convicted.³³ Warrantless arrests and searches incident to them were not the norm.³⁴ Thus, search incident to arrest, as used in the era of the Framers, provided today’s law enforcement scant guidance.

At first, even less direction was provided by Supreme Court precedent. The Court first acknowledged search incident to arrest in *Weeks v. United States*, where it openly conceded that such a doctrine was not relevant to the case.³⁵

27. Logan, *supra* note 23, at 385.

28. *Id.*

29. Thomas Y. Davies, *Recovering the Original Fourth Amendment*, 98 MICH. L. REV. 547, 577 (1999).

30. See *id.* Davies quotes Blackstone’s declaration that “considering what manner of men are for the most part put upon these offices, it is perhaps very well that they are generally kept in ignorance [of the full extent of the authority of their office].” *Id.* at 577 n.69 (alteration in original) (citation omitted).

31. *Id.* at 578.

32. *Id.* at 577-78.

33. See Logan, *supra* note 23, at 385-86.

34. *Id.* (“Thus, as a practical matter, authorities had relatively little occasion to arrest persons in the absence of a warrant, and as a result had only limited recourse to conduct searches incident to arrest . . .”).

35. *Weeks v. United States*, 232 U.S. 383, 392 (1914), overruled by *Mapp v. Ohio*, 367 U.S. 643 (1961). Justice Day, writing for the Court, declared,

What, then, is the present case? Before answering that inquiry specifically, it may be well by a process of exclusion to state what it is not. It is not an assertion of the right on the part of the government always recognized under English and American law, to

Thus, the first mention of search incident to arrest in a Court opinion was mere dictum. Further, the scope of the search right *Weeks* articulated was quite narrow. The Court made no mention of the right to search for weapons, referring only to the recovery of “fruits or evidences of crime.”³⁶ Further, *Weeks* made no reference to any area beyond the arrestee’s body, articulating only the right to “search the person of the accused.”³⁷

As Justice Frankfurter alluded,³⁸ the Court then “elevated [Weeks] to a decision” in *Carroll v. United States* by intoning, “When a man is legally arrested for an offense, whatever is found upon his person or in his control which it is unlawful for him to have and which may be used to prove the offense may be seized and held as evidence in the prosecution.”³⁹ By mentioning an area “in his [the arrestee’s] control,”⁴⁰ *Carroll* outstripped *Weeks* by extending the scope of search incident to arrest beyond the arrestee’s person. Furthermore, *Carroll* performed this legerdemain in yet more dictum, for the Court explained that, in its case, “[t]he right to search and the validity of the seizure are not dependent on the right to arrest.”⁴¹

The dicta in *Weeks* and *Carroll* emboldened the Court in *Agnello v. United States* to broaden search incident to arrest from an arrestee’s “control” to include “the place where the arrest is made”⁴² or “one’s house.”⁴³ *Agnello* gave no further precision because the Court, noting that the relevant search was of a house “several blocks distant” from the arrest, simply ruled that search incident to arrest “[did] not extend to other places.”⁴⁴ *Agnello* also made a change regarding the object of search incident to arrest; the Court expanded the list of items officers could search for to include “weapons and other things to effect an escape from custody.”⁴⁵ As in the prior cases, the Court failed to make its expansion of items subject to search part of its holding, for the search of Agnello’s home was found to be outside the scope of the rule.⁴⁶

The Court maintained an expansive view of search incident to arrest in its

search the person of the accused when legally arrested, to discover and seize the fruits or evidences of crime.

Id.

36. *Id.*

37. *Id.*

38. *United States v. Rabinowitz*, 339 U.S. 56, 75 (1950) (Frankfurter, J., dissenting), overruled in part by *Chimel v. California*, 395 U.S. 752 (1969).

39. *Carroll v. United States*, 267 U.S. 132, 158 (1925) (citations omitted).

40. *Id.*

41. *Id.* Instead, the *Carroll* Court based the lawfulness of the search on the automobile exception, a rule it created. *See id.* at 158-59, 162.

42. *Agnello v. United States*, 269 U.S. 20, 30 (1925).

43. *See id.* at 32.

44. *Id.* at 30-31.

45. *Id.* at 30.

46. *Agnello* ruled that the search of Agnello’s home “[could not] be sustained as an incident of the arrests.” *Id.* at 31.

next case, *Marron v. United States*.⁴⁷ In *Marron*, federal prohibition agents arrested Birdsall—a co-conspirator of Marron’s—because he was operating a speakeasy, and thus recovered some utility bills and a ledger listing inventories of liquor.⁴⁸ The Court determined that the officers had a right to search “the place” of arrest and that such right “extended to all parts of the premises used for the unlawful purpose,”⁴⁹ even though the premises were large enough to have “six or seven rooms.”⁵⁰ Yet the *Marron* Court, as with previous search incident to arrest case law, failed to link its broad language with the facts of the case. While expansively describing search incident to arrest’s scope, it justified the search that actually occurred in the case by noting the nearness of the items to the arrestee.⁵¹ The ledger was in the arrestee Birdsall’s “immediate possession and control.”⁵²

Thus, a rule confidently and repeatedly announced in Court dicta did indeed ripen into a decision. This evolution, however, was not without cost. The rule’s scope was not closely tied to the facts of the cases in which it was declared. Search incident to arrest’s actual boundaries were not supported by a fully explained or understood rationale and therefore were vulnerable to alteration and even retrenchment.

B. Constant Inconsistency

Less than five years after *Marron*, the Court began a dramatic contraction of search incident to arrest in *Go-Bart Importing Co. v. United States*.⁵³ Although *Go-Bart*, like *Marron*, was a prohibition case, it presented quite different police behavior.⁵⁴ The agents falsely claimed to have a search warrant, took the arrestee’s keys, and “by threat of force compelled him to open a desk and safe.”⁵⁵ The *Go-Bart* Court found the search unreasonable, condemning the officers’ actions as making a “general and apparently unlimited search, ransacking the desk, safe, filing cases and other parts of the office.”⁵⁶ Further, the Court deemed the search of an office to be “a general exploratory search in the hope that evidence of crime might be found.”⁵⁷ Condemning the search of a single office seemed inconsistent with *Marron*’s generous rule. Rather than question the boundaries previously allowed by *Marron*, the *Go-Bart* Court merely

47. 275 U.S. 192 (1927).

48. *Id.* at 194.

49. *Id.* at 199.

50. *Id.* at 193-94.

51. *See id.* at 199.

52. *Id.*

53. 282 U.S. 344 (1931).

54. *See id.* at 349-50.

55. *Id.* at 349.

56. *Id.* at 358.

57. *Id.* (citation omitted).

distinguished its facts.⁵⁸ *Marron* simply lacked the reprehensible behavior of the *Go-Bart* agents, for the officers in *Marron* executed a valid warrant without threatening force or rummaging the place in a general search.⁵⁹ Further, *Go-Bart* noted that *Marron*'s officers collected items that "were visible and accessible and in the offender's immediate custody."⁶⁰

The Court continued its contraction of search incident to arrest in *United States v. Lefkowitz*,⁶¹ still another prohibition case. *Lefkowitz* determined that the search of a "ten feet wide and twenty feet long"⁶² room violated the Fourth Amendment.⁶³ As did *Go-Bart*, *Lefkowitz* distinguished *Marron* as a case where the evidence was seized "in plain view" and thus collected without a search being performed.⁶⁴ Such reasoning seemed forced. However much *Lefkowitz* attempted to distinguish *Marron* rather than simply overturn it, the fact remained that in the space of five years, search incident to arrest had shrunk from supporting a search of an entire home to failing to justify the search of one room.⁶⁵

Lefkowitz's severe restriction of search incident to arrest's scope was followed by yet another expansion of the rule in *Harris v. United States*.⁶⁶ Whereas *Lefkowitz* deemed the search of "room 604"⁶⁷ to be "unrestrained,"⁶⁸ *Harris* found a "careful and thorough"⁶⁹ five-hour search of an entire apartment to be "basically reasonable."⁷⁰ The search in *Harris* was so intrusive that it included lifting carpets, stripping bed linen, turning over a mattress, and opening a sealed envelope in a bureau drawer.⁷¹ The Court reasoned that since Harris "was in exclusive possession of a four room apartment," the arrestee's control "extended quite as much to the bedroom in which the . . . [evidence was] found as to the living room in which he was arrested."⁷² The dissent in *Harris* recognized the glaring incongruity between the Court's ruling and recent precedent, declaring that finding the search lawful in *Harris* was tantamount to throwing *Go-Bart* and *Lefkowitz* "to the winds."⁷³

58. *Id.* ("Plainly the case before us is essentially different from *Marron v. United States*.").

59. *Id.*

60. *Id.*

61. 285 U.S. 452 (1932).

62. *Id.* at 458.

63. *Id.* at 467.

64. *Id.* at 465.

65. The Court's delicacy in refusing to openly overturn *Marron* was somewhat ironic in light of the fact that this precedent had its origin in *dictum* in *Weeks* and *Carroll*.

66. 331 U.S. 145 (1947), *overruled in part by* *Chimel v. California*, 395 U.S. 752 (1969).

67. *Lefkowitz*, 285 U.S. at 458.

68. *Id.* at 464.

69. *Harris*, 331 U.S. at 149.

70. *See id.* at 155.

71. *Id.* at 169 (Frankfurter, J., dissenting).

72. *Id.* at 152 (majority opinion).

73. *Id.* at 167 (Frankfurter, J., dissenting).

One year after *Harris*, the Court abruptly reversed course in *Trupiano v. United States*.⁷⁴ In *Trupiano*, agents raided a farm and seized an illegal still without troubling themselves to obtain a search warrant.⁷⁵ The Court found the seizure improper,⁷⁶ denying officers the use of search incident to arrest because “[t]he mere fact that there is a valid arrest does not ipso facto legalize a search or seizure without a warrant.”⁷⁷ The Court deemed search incident to arrest such a “strictly limited right” that it “[grew] out of the inherent necessities of the situation.”⁷⁸ Yet what those “necessities” were was far from clear. The Court groped to explain by offering that “there must be something more in the way of necessity than merely a lawful arrest” and that “there must be some other factor in the situation that would make it unreasonable or impracticable to require the arresting officer to equip himself with a search warrant.”⁷⁹ The facts in the case presented no such necessities, and the Court gave no examples to clarify its ruling.

Outside of a half-hearted attempt to distinguish *Harris* as involving the unexpected discovery of evidence for which officials could not have thought to seek a warrant before the arrest (while *Trupiano*'s officials, by comparison, had plenty of forewarning of the need for a warrant), *Trupiano* did not even pretend to reconcile its ruling with *Harris*.⁸⁰ Indeed, *Trupiano* admitted, “These factual differences may or may not be of significance so far as general principles are concerned.”⁸¹ The *Trupiano* Court flatly declared, “We do not take occasion here to reexamine the situation involved in [Harris].”⁸²

The Court lurched back to broadening search incident to arrest only two years later in *United States v. Rabinowitz*, a case involving an hour and a half search of a one-room office.⁸³ The Court in *Rabinowitz* found the search to be “incident to a lawful arrest . . . and therefore valid.”⁸⁴ Although refusing to bind itself to any “ready litmus paper test,” the Court did specify five factors which pointed toward reasonableness.⁸⁵ Some elements were hardly novel, such as the fact that the search and seizure were based on a valid arrest and that possession of the recovered evidence (forged stamps) was a crime.⁸⁶ Two factors—that “the

74. 334 U.S. 699 (1948), *overruled in part* by *United States v. Rabinowitz*, 339 U.S. 56 (1950).

75. *Id.* at 701-03.

76. *Id.* at 710.

77. *Id.* at 708.

78. *Id.*

79. *Id.*

80. *Id.* at 708-09.

81. *Id.* at 709.

82. *Id.* at 708.

83. *United States v. Rabinowitz*, 339 U.S. 56, 58-59, 63-64 (1950), *overruled in part* by *Chimel v. California*, 395 U.S. 752 (1969).

84. *Id.* at 63.

85. *Id.* at 63-64.

86. *Id.* at 64.

room was small and under the immediate and complete control of [the arrestee]” and that “the search did not extend beyond the room used for unlawful purposes”⁸⁷—appeared to be reformulations of norms stated in prior cases. One factor—that “the place of the search was a business room to which the public, including the officers, was invited”⁸⁸—seemed to be a new creation crafted for the case at hand. *Rabinowitz* did not explain the origins of the factors. Nor did the Court seem to place much reliance on them, for it declared that “[w]hat is a reasonable search is not to be determined by any fixed formula.”⁸⁹

In dissent, Justice Frankfurter vehemently disagreed with the ruling, essentially branding the majority’s reasoning a “distortion”⁹⁰ and a “farce.”⁹¹ He emphasized that the “basic roots” of search incident to arrest “lie in necessity.”⁹² The two necessities were: “first, in order to protect the arresting officer and to deprive the prisoner of potential means of escape, and secondly, to avoid destruction of evidence by the arrested person.”⁹³ These two imperatives, though not appreciated at the time, would become guideposts for future search incident to arrest case law.

C. An Attempt to Establish a Rational and Lasting Rule

The Court itself knew of the uncertainty plaguing search incident to arrest.⁹⁴ In *Chimel v. California*, Justice Stewart, who wrote the Court’s opinion, compared the precedent of this doctrine to a swinging pendulum⁹⁵ that was “hardly founded on an unimpeachable line of authority.”⁹⁶ *Chimel* therefore endeavored to bring consistency to this area of law by anchoring it to specifics, averring:

As Mr. Justice Frankfurter put it: “To say that the search must be reasonable is to require some criterion of reason. It is no guide at all either for a jury or for district judges or the police to say that an ‘unreasonable search’ is forbidden—that the search must be reasonable.

What is the test of reason which makes a search unreasonable?”⁹⁷

In this regard, Justice Stewart wondered if it were reasonable to search a person’s

87. *Id.*

88. *Id.*

89. *Id.* at 63.

90. *Id.* at 71 (Frankfurter, J., dissenting).

91. *Id.* at 72.

92. *Id.*

93. *Id.* (internal citation omitted).

94. See *Chimel v. California*, 395 U.S. 752, 755 (1969) (“The decisions of this Court bearing upon that question [search incident to arrest] have been far from consistent, as even the most cursory review makes evident.”).

95. *Id.* at 758.

96. *Id.* at 760.

97. *Id.* at 765 (quoting *Rabinowitz*, 399 U.S. at 83 (Frankfurter, J., dissenting)).

home simply because he or she was arrested in it, what made the same search unreasonable when the arrest occurred out on the front lawn or just down the street?⁹⁸ “Under such an unconfined analysis, Fourth Amendment protection in this area would approach the evaporation point.”⁹⁹

Chimel therefore set out to craft a rule that was grounded in the practicalities of arrests. One such concrete concern involved officer safety, for “[w]hen an arrest is made, it is reasonable for the arresting officer to search the person arrested in order to remove any weapons that the latter might seek to use in order to resist arrest or effect his escape.”¹⁰⁰ Equally valid was the interest in protecting the case. *Chimel* explained, “In addition, it is entirely reasonable for the arresting officer to search for and seize any evidence on the arrestee’s person in order to prevent its concealment or destruction.”¹⁰¹ The actual necessities of the arrest setting thus helped identify the objects for which officers could reasonably search.

Practical consequences of arrests also informed the *Chimel* Court in its determination of search incident to arrest’s scope. Justice Stewart reasoned:

[T]he area into which an arrestee might reach in order to grab a weapon or evidentiary items must, of course, be governed by a like rule. A gun on a table or in a drawer in front of one who is arrested can be as dangerous to the arresting officer as one concealed in the clothing of the person arrested.¹⁰²

There was thus “ample justification . . . [to allow] a search of the arrestee’s person and the area ‘within his immediate control’—construing that phrase to mean the area from within which he might gain possession of a weapon or destructible evidence.”¹⁰³

Chimel made clear that permitting officers to move beyond these common sense boundaries would destroy any effort at meaningfully limiting searches incident to arrest. Justice Stewart warned, “No consideration relevant to the Fourth Amendment suggests any point of rational limitation, once the search is allowed to go beyond the area from which the person arrested might obtain weapons or evidentiary items.”¹⁰⁴

98. *Id.* at 764-65.

99. *Id.* at 765.

100. *Id.* at 762-63. Justice Stewart explained, “Otherwise, the officer’s safety might well be endangered, and the arrest itself frustrated.” *Id.* at 763.

101. *Id.*

102. *Id.*

103. *Id.*

104. *Id.* at 766. In this regard, Justice Stewart quoted from Justice Jackson’s dissent in *Harris*:
The difficulty with this problem for me is that once the search is allowed to go beyond the person arrested and the objects upon him or in his immediate physical control, I see no practical limit short of that set in the opinion of the Court—and that means to me no limit at all.

Id. at 766 n.11 (citation omitted).

D. Insinuation of Per Se Search Rights into Chimel's Scope Rule

In defending its newly formed rules, *Chimel* explicitly rejected the dissent's contention that "so long as there is probable cause to search the place where an arrest occurs, a search of that place should be permitted even though no search warrant has been obtained," because such an argument would create a per se right to search the room of arrest.¹⁰⁵ Justice Stewart responded, "[W]e can see no reason why, simply because some interference with an individual's privacy and freedom of movement has lawfully taken place, further intrusions should automatically be allowed despite the absence of a warrant that the Fourth Amendment would otherwise require."¹⁰⁶

Despite such an effort, absolute search rights seeped into search incident to arrest case law. One of the most significant examples occurred in *United States v. Robinson*,¹⁰⁷ a case involving an arrest for driving on a revoked license. When the arresting officer patted down Robinson's breast pocket, he felt something, but he "couldn't tell what it was."¹⁰⁸ The officer then recovered a "crumpled up cigarette package," finding heroin inside.¹⁰⁹

Justice Rehnquist, the author of the Court's opinion in *Robinson*, took a categorical approach to search incident to arrest. He perceived arrests, as a kind of seizure, to present peril to police, for it was "scarcely open to doubt that the . . . extended exposure which follows the taking of a suspect into custody" created "far greater" danger than simple street encounters.¹¹⁰ He thus refused to tailor the intrusiveness of the search on the basis of the severity of the crime committed, instead choosing to treat "all custodial arrests alike for purposes of search justification."¹¹¹

The same one-size-fits-all analysis was applied to measuring the scope of the search permitted by an arrest:

A police officer's determination as to how and where to search the person of a suspect whom he has arrested is necessarily a quick *ad hoc* judgment which the Fourth Amendment does not require to be broken

105. *Id.* at 766 n.12.

106. *Id.*

107. 414 U.S. 218 (1973).

108. *Id.* at 223.

109. *Id.*

110. *Id.* at 234-35.

111. See *id.* at 235. Justice Rehnquist declared,

Nor are [we] inclined, on the basis of what seems to us to be a rather speculative judgment, to qualify the breadth of the general authority to search incident to a lawful custodial arrest on an assumption that persons arrested for the offense of driving while their licenses have been revoked are less likely to possess dangerous weapons than are those arrested for other crimes.

Id. at 234.

down in each instance into an analysis of each step of the search.¹¹²

Even though the purposes of search incident to arrest were the officer's "need to disarm and to discover evidence," the right to search in the individual case did not depend on what a judge might "later decide was the probability in a particular arrest situation that weapons or evidence would in fact be found upon the person of the suspect."¹¹³ In fact, the officer's failure in this case to have any fear of Robinson or to suspect he was armed was simply "of no moment."¹¹⁴ Thus, the mere existence of a lawful arrest enabled an officer in every case to search for both weapons and evidence.¹¹⁵ Further, *Robinson* promoted this absolute version of search incident to arrest as not merely an exception to the warrant requirement, but as an "affirmative authority to search" meeting the Fourth Amendment's reasonableness requirement.¹¹⁶ *Robinson*'s laissez-faire attitude toward police marked a dramatic retreat from *Chimel*'s insistence that a warrantless search be based on "the inherent necessities of the situation"¹¹⁷ and that the warrant requirement "is not lightly to be dispensed with, and 'the burden is on those seeking (an) exemption (from the requirement) to show the need for it.'"¹¹⁸

Another case that crafted an absolute search right for officers performing arrests was actually ambivalent about creating a *per se* rule.¹¹⁹ In *Maryland v. Buie*, police lawfully entered Buie's home to arrest him for armed robbery.¹²⁰ When an officer called down into the basement, Buie emerged with his hands up and was arrested.¹²¹ Afterward, a detective went down into the basement "in case there was someone else down there" and noticed a red running suit worn by the robber in plain view.¹²² Buie's motion to suppress the running suit was denied.¹²³

In considering whether the detective's entry into the basement after the defendant's exit and arrest was reasonable, the Court, in an opinion written by

112. *Id.* at 235.

113. *Id.*

114. *Id.* at 236.

115. *See id.* at 235-36.

116. *Id.* at 226, 235.

117. *Chimel v. California*, 395 U.S. 752, 759 (1969) (quoting *Trupiano v. United States*, 344 U.S. 699, 708 (1948)). *Chimel* declared that a warrant could be excused only with a "showing by those who seek exemption from the constitutional mandate that the exigencies of the situation made that course imperative." *Id.* at 761 (quoting *McDonald v. United States*, 335 U.S. 451, 455-56 (1948)).

118. *Id.* at 762 (alteration in original) (quoting *United States v. Jeffers*, 342 U.S. 48, 51 (1951)).

119. *Maryland v. Buie*, 494 U.S. 325, 336-37 (1990).

120. *Id.* at 328.

121. *Id.*

122. *Id.* (internal citation omitted).

123. *Id.*

Justice White, delineated two kinds of “protective sweep[s].”¹²⁴ *Buie* specified that for “closets and other spaces immediately adjoining the place of arrest from which an attack could be immediately launched,” officers were permitted to look “as a precautionary matter and without probable cause or reasonable suspicion.”¹²⁵ In contrast, for officers to venture beyond this area, *Buie* held that “there must be articulable facts which, taken together with the rational inferences from those facts, would warrant a reasonably prudent officer in believing that the area to be swept harbors an individual posing a danger to those on the arrest scene.”¹²⁶ Since these searches are meant to protect arresting officers, the sweeps must amount to a “cursory inspection of those spaces where a person may be found,”¹²⁷ lasting “no longer than is necessary to dispel reasonable suspicion of danger and in any event no longer than it takes to complete the arrest and depart the premises.”¹²⁸

In requiring some level of objective justification for the second kind of sweep (covering the entire home), *Buie* avoided a *per se* search right in favor of a rule providing power “no more than necessary to protect the officer from harm.”¹²⁹ For support, Justice White noted the Court’s earlier refusal in *Terry v. Ohio* to create a bright-line rule when crafting a right to frisk a person on the street.¹³⁰ *Buie* recognized that

despite the danger that inheres in on-the-street encounters and the need for police to act quickly for their own safety, the Court in *Terry* did not adopt a bright-line rule authorizing frisks for weapons in all confrontational encounters. Even in high crime areas, where the possibility that any given individual is armed is significant, *Terry* requires reasonable, individualized suspicion before a frisk for weapons can be conducted. That approach is applied to the protective sweep of a house.¹³¹

This explicit adherence to an individualized suspicion requirement for protective sweeps of homes might have obscured the fact that *Buie* created an absolute right for officers to search “spaces immediately adjoining the place of arrest.”¹³² Unlike the protective sweep of the home, *Buie* extended search incident to arrest authority to include looking for persons in these areas without any additional

124. *Id.* at 327, 334-35. In this regard, the Court posed the issue as requiring that it “decide what level of justification is required . . . before police officers, while effecting the arrest of a suspect in his home pursuant to an arrest warrant, may conduct a warrantless protective sweep of all or part of the premises.” *Id.* at 327 (emphasis added).

125. *Id.* at 334.

126. *Id.*

127. *Id.* at 335.

128. *Id.* at 335-36 (citation omitted).

129. *Id.* at 334.

130. *Id.* at 334 n.2 (citing *Terry v. Ohio*, 392 U.S. 1 (1968)).

131. *Id.*

132. *Id.* at 334.

justification.¹³³

The *Buie* Court determined that protective sweeps were not constrained by *Chimel*'s prohibition against extending the search beyond an arrestee's person or area of immediate control because the two cases were simply different.¹³⁴ *Chimel* concerned the prospect of extending search incident to arrest to the entire house, while *Buie* merely involved a more limited protective sweep of the home.¹³⁵ Moreover, *Chimel*'s focus of danger was the arrestee; in *Buie*, the peril came from "unseen third parties."¹³⁶ *Buie*, therefore, established an absolute right to search in all search incident to arrest precedent without viewing its extension as undermining the integrity of *Chimel*'s limitations.

E. Search Incident to Arrest and Vehicles

Perhaps the most dramatic expansion of search incident to arrest occurred in *New York v. Belton*,¹³⁷ where the Court devised a bright-line boundary for police searches of vehicles. In *Belton*, a police officer, upon pulling over four men in a speeding car, smelled marijuana and observed an item of marijuana paraphernalia on the floor of the vehicle.¹³⁸ After patting down the occupants and splitting them up from each other on the road "so they would not be in physical touching area of each other," the officer searched the passenger compartment of the car.¹³⁹ As a result, he unzipped a pocket of a black leather jacket on the back seat and found cocaine.¹⁴⁰

Belton was written by Justice Stewart, the same author who had taken such care in crafting a reasoned basis for the scope of search incident to arrest in *Chimel*.¹⁴¹ In *Belton*, Justice Stewart's concern was the workability of *Chimel*'s rule when applied to cars.¹⁴² The *Belton* Court worried that Fourth Amendment protections could "only be realized if the police are acting under a set of rules which, in most instances, makes it possible to reach a correct determination beforehand as to whether an invasion of privacy is justified in the interest of law enforcement."¹⁴³ Justice Stewart noted that the Fourth Amendment was "primarily intended to regulate the police in their day-to-day activities and thus ought to be expressed in terms that are readily applicable by the police in the context of the law enforcement activities in which they are necessarily

133. *See id.*

134. *See id.* at 336 (citing *Chimel v. California*, 395 U.S. 752 (1969)).

135. *Id.*; *Chimel*, 395 U.S. at 754.

136. *Buie*, 494 U.S. at 336; *Chimel*, 395 U.S. at 766-67.

137. 453 U.S. 454 (1981).

138. *Id.* at 455-56. The officer observed a "Supergold" wrapper, which he associated with marijuana, on the car's floor. *Id.*

139. *Id.* at 456.

140. *Id.*

141. *Id.* at 454-55.

142. *See id.* at 458-59.

143. *Id.* at 458 (citation omitted).

engaged.”¹⁴⁴ If search incident to arrest presented too complex a rule, neither citizens nor police would be able to know the scope of their protection against unreasonable search and seizure.¹⁴⁵

Belton found that lower courts lacked a “workable definition” of *Chimel*’s “area within the immediate control of the arrestee” when police arrested motorists.¹⁴⁶ The Court therefore attempted to draw a simple bright line by holding that “when a policeman has made a lawful custodial arrest of the occupant of an automobile, he may, as a contemporaneous incident of that arrest, search the passenger compartment of that automobile.”¹⁴⁷ Thus, anything inside the passenger compartment—including containers—could be searched under *Belton*, even if such containers “could hold neither a weapon nor evidence of the criminal conduct for which the suspect was arrested.”¹⁴⁸

Even though *Belton* now enabled police to search areas “not inevitably” within an arrestee’s control, and in containers that could not hold a weapon or evidence, Justice Stewart did not view this creation of a *per se* search right as an expansion of *Chimel*’s carefully considered boundaries.¹⁴⁹ He stated, perhaps more out of hope than of conviction, that “[o]ur holding today does no more than determine the meaning of *Chimel*’s principles in this particular and problematic context. It in no way alters the fundamental principles established in the *Chimel* case regarding the basic scope of searches incident to lawful custodial arrests.”¹⁵⁰

This pretense became harder to maintain in the Court’s next case, *Thornton v. United States*,¹⁵¹ in which a driver stopped and exited his vehicle before the police officer following him had a chance to pull him over. Chief Justice Rehnquist, who wrote the Court’s opinion, asserted that “[i]n all relevant aspects, the arrest of a suspect who is next to a vehicle presents identical concerns regarding officer safety and the destruction of evidence as the arrest of one who is inside the vehicle.”¹⁵² The Court found the stress of the arrest “no less merely because the arrestee exited his car before the officer initiated contact” and further

144. *Id.* (citation omitted). The Court continued as follows:

A highly sophisticated set of rules, qualified by all sorts of ifs, ands, and buts and requiring the drawing of subtle nuances and hairline distinctions, may be the sort of heady stuff upon which the facile minds of lawyers and judges eagerly feed, but they may be “literally impossible of application by the officer in the field.”

Id. (citation omitted).

145. *Id.* at 459-60. *Belton* declared that “[w]hen a person cannot know how a court will apply a settled principle to a recurring factual situation, that person cannot know the scope of his constitutional protection, nor can a policeman know the scope of his authority.” *Id.*

146. *Id.* at 460. The *Belton* Court specifically noted the confusion when the area at issue “arguably include[d] the interior of an automobile and the arrestee is its recent occupant.” *Id.*

147. *Id.* (internal footnote omitted).

148. *Id.* at 461.

149. *Id.* at 460 n.3.

150. *Id.*; see generally *Chimel v. California*, 395 U.S. 752 (1969).

151. 541 U.S. 615 (2004).

152. *Id.* at 621.

concluded that the arrestee was “[no] less likely to attempt to lunge for a weapon or to destroy evidence if he is outside of, but still in control of, the vehicle.”¹⁵³ *Thornton* therefore extended *Belton* to include those encounters “when an officer does not make contact until the person arrested has left the vehicle.”¹⁵⁴

To support this expansion, Chief Justice Rehnquist emphasized the “need for a clear rule” to avoid the “ad hoc determinations on the part of officers in the field and reviewing courts.”¹⁵⁵ *Thornton* therefore sought to preserve its bright-line analysis despite the fact that “not all contraband in the passenger compartment is likely to be readily accessible to a ‘recent occupant.’”¹⁵⁶ Such an improbability, in fact, existed in *Thornton* itself because “[i]t . . . [was] unlikely in this case that . . . [Thornton] could have reached under the driver’s seat for his gun once he was outside of his automobile.”¹⁵⁷ Thus, the *Thornton* Court, in explaining its concern about forcing officials to apply a case-by-case test to determine the legality of a particular search, readily applied such an ad hoc analysis in its own case.

In the wake of *Chimel*’s effort to establish a reasoned basis for the boundaries of search incident to arrest, the Court increasingly lost faith that officers in the field or judges in the courtroom would be able to accurately and consistently apply its rule. To resolve this doubt, *Thornton* created ever-broadening bright-line rules meant to enable even the dimmest officials to reach the proper conclusion about where to search incident to arrest. The *per se* search rights, however, tended to undermine the very reasoning *Chimel* had established in the first place in order to prevent pendulum swings in this area of the law. Thus, on the threshold of *Arizona v. Gant*, the Court’s search incident to arrest doctrine for vehicles had two incompatible aims—to adhere to a rule strictly limited to “the inherent necessities of the situation,”¹⁵⁸ while at the same time offering a “sort of generalization” for ready understanding by officers in the field.¹⁵⁹

II. *ARIZONA V. GANT*

A. *The Facts*

On August 25, 1999, Tucson police officers Griffith and Reed visited a home at 2524 North Walnut Avenue to check out an anonymous tip that the house was being used to sell drugs.¹⁶⁰ When Rodney Gant answered the door, the officers

153. *Id.*

154. *Id.* at 617.

155. *Id.* at 623.

156. *Id.* at 622.

157. *Id.*

158. *Chimel v. California*, 395 U.S. 752, 759 (1969).

159. *Thornton*, 541 U.S. at 623.

160. *Arizona v. Gant*, 129 S. Ct. 1710, 1714 (2009).

asked to speak to the owner of the house.¹⁶¹ Gant identified himself and told the officers that the owner was not at home but would return later that afternoon.¹⁶² The officers later conducted a records check on Gant, learning that he had an outstanding arrest warrant for driving with a suspended license.¹⁶³ That evening, the officers returned to the home and found “a man near the back of the house and a woman in a car parked in front of it.”¹⁶⁴ Upon arrival of a third officer, the police arrested the man “for providing a false name and the woman for possessing drug paraphernalia.”¹⁶⁵ After the two arrestees were handcuffed and placed separately in the two patrol cars, Gant pulled into the driveway.¹⁶⁶ The officers recognized his vehicle, shining a flashlight into the car to confirm his identity.¹⁶⁷ After Gant parked his car at the end of the driveway, exited, and shut his door, Officer Griffith—who was about thirty feet away—called to him.¹⁶⁸ Approaching each other, the two met ten to twelve feet from Gant’s car, where Griffith immediately arrested and handcuffed Gant.¹⁶⁹ Griffith then called for backup, and when two more officers arrived, the police locked Gant in the back of the newly available vehicle.¹⁷⁰ Within minutes of the arrest, the officers searched Gant’s car and found a gun and a bag of cocaine in the pocket of a jacket on the backseat.¹⁷¹ Gant later moved to suppress the evidence found in his car as obtained in violation of the Fourth Amendment.¹⁷²

B. The Court’s Opinion

The *Gant* Court, in an opinion written by Justice Stevens, established a new two-part rule for search of an automobile incident to arrest.¹⁷³ *Gant* first held that police could “search a vehicle incident to a recent occupant’s arrest only when the arrestee is unsecured and within reaching distance of the passenger compartment at the time.”¹⁷⁴ The Court next concluded that “circumstances unique to the vehicle context justify a search incident to arrest when it is ‘reasonable to believe evidence relevant to the crime of arrest might be found in

161. *Id.*

162. *Id.* at 1714-15.

163. *Id.* at 1715.

164. *Id.*

165. *Id.*

166. *Id.*

167. *Id.*

168. *Id.*

169. *Id.*

170. *Id.*

171. *Id.*

172. *Id.*

173. See *id.* at 1714.

174. *Id.* at 1719. *Gant* offered the converse of the same rule, holding that “*Belton* does not authorize a vehicle search incident to a recent occupant’s arrest after the arrestee has been secured and cannot access the interior of the vehicle.” *Id.* at 1714.

the vehicle.”¹⁷⁵ These two rules are potentially in conflict: the first restricts search incident to arrest to its original necessities (searching an area only because the arrestee might reach into it for a weapon or evidence), while the second expands the rule, providing police with an entirely new and independent basis (reason to believe evidence of the offense exists) for searching a vehicle incident to arrest.¹⁷⁶

The incompatibility of *Gant*'s “reaching distance” limit with its “reason to believe” or “offense-related evidence” expansion can be gleaned by the differing treatment each rule was given in the opinion. The rationale justifying the reaching distance limit required most of the opinion, while the support for “offense-related evidence” fit within a single paragraph.¹⁷⁷ The very structure of the opinion, therefore, hinted that Justice Stevens's primary aim in *Gant* was to place a limit on the recent expansions of search incident to arrest's scope. The “offense-related evidence” rule seemed grafted on in an attempt to gain the needed fifth vote of Justice Scalia. In the one paragraph devoted to explaining the “offense-related evidence” expansion, the *Gant* Court mentioned Justice Scalia's concurring opinion in *Thornton*, urging its implementation.¹⁷⁸ Further, Justice Scalia himself made no secret of his need to hold his nose in order to concur with the majority, explaining:

It seems to me unacceptable for the Court to come forth with a 4-to-1-to-4 opinion that leaves the governing rule uncertain. I am therefore confronted with the choice of either leaving the current understanding of *Belton* and *Thornton* in effect, or acceding to what seems to me the artificial narrowing of those cases adopted by Justice Stevens. The latter, as I have said, does not provide the degree of certainty I think desirable in this field; but the former opens the field to what I think are plainly unconstitutional searches—which is the greater evil. I therefore join the opinion of the Court.¹⁷⁹

For its primary goal—the “reaching distance” limit on *Belton*—*Gant* took great care in laying a proper foundation, returning to the original understanding of search incident to arrest as merely a “well-delineated” exception to the warrant requirement.¹⁸⁰ This characterization was important, for it represented a significant break with the bolder claim of the early bright line case of *Robinson*, which saw search incident to arrest as reasonable in its own right.¹⁸¹ Since search

175. *Id.* at 1719 (quoting *Thornton v. United States*, 541 U.S. 615, 632 (2004)).

176. *See id.* at 1714.

177. *See id.* at 1719.

178. *See id.* at 1718.

179. *Id.* at 1725 (Scalia, J., concurring).

180. *Id.* at 1716 (majority opinion).

181. *See United States v. Robinson*, 414 U.S. 218, 226 (1973). In *Robinson*, the Court gave search incident to arrest an independent basis of legitimacy, averring, “Since the statements in the cases speak not simply in terms of an exception to the warrant requirement, but in terms of an affirmative authority to search, they clearly imply that such searches also meet the Fourth

incident to arrest was an exception to a general rule, deviating from the norm of procuring a warrant rested on the interests of “officer safety and evidence preservation.”¹⁸² These very interests, in turn, created a limit which ensured that “the scope of a search incident to arrest . . . [was] commensurate with its purposes of protecting arresting officers and safeguarding any evidence of the offense that an arrestee might conceal or destroy.”¹⁸³ If search incident to arrest were concerned with the arrestee’s access to weapons or evidence, then the only area of worry would be the area from which he “might gain possession.”¹⁸⁴ Thus, “[i]f there is no possibility that an arrestee could reach into the area that law enforcement officers seek to search, both justifications for the search-incident-to-arrest exception are absent and the rule does not apply.”¹⁸⁵

Gant then considered the weapons and evidence interests involved in *Belton*.¹⁸⁶ Interestingly, in analyzing a case that established a bright-line rule, *Gant* deemed *Belton*’s specific facts to be quite relevant. *Belton* had applied *Chimel* in a context involving a “lone police officer” confronting four arrestees, none of whom were handcuffed.¹⁸⁷ By framing *Belton* within these facts, *Gant* attempted to limit *Belton* to the items in a passenger compartment that are generally within the arrestee’s reaching distance.¹⁸⁸ Justice Stevens then lamented that the Court’s opinion “has been widely understood to allow a vehicle search incident to the arrest of a recent occupant even if there is no possibility the arrestee could gain access to the vehicle at the time of the search.”¹⁸⁹ *Gant* blamed this misunderstanding on Justice Brennan’s dissent in *Belton*, which warned that *Belton*’s ruling would mean that “the interior of a car is always within the immediate control of the arrestee who has recently been in the car.”¹⁹⁰ Scapegoating Justice Brennan enabled the *Gant* Court to conveniently forget its own support for a broad interpretation of *Belton* in *Thornton*, where the Court upheld the search even though the officer had “handcuffed . . . [Thornton], informed him that he was under arrest, and placed him in the back seat of the patrol car” before searching the vehicle.¹⁹¹

Gant then applied its “reaching distance” rule, determining that there was no

Amendment’s requirement of reasonableness.” *Id.*

182. *Gant*, 129 S. Ct. at 1716 (citation omitted).

183. *Id.*

184. *Id.*

185. *Id.*

186. *Id.* at 1716-17.

187. *Id.* *Gant* noted that “[t]he officer [in *Belton*] was unable to handcuff the occupants because he had only one set of handcuffs.” *Id.* at 1717 n.1.

188. *Gant* noted that *Belton*’s “holding was based in large part on our assumption ‘that articles inside the relatively narrow compass of the passenger compartment of an automobile are in fact generally, even if not inevitably, within the area into which an arrestee might reach.’” *Id.* at 1717 (quoting *New York v. Belton*, 453 U.S. 454, 460 (1981)).

189. *Id.* at 1718.

190. *Id.* (emphasis in original) (quoting *Belton*, 453 U.S. at 466 (Brennan, J., dissenting)).

191. *Thornton v. United States*, 541 U.S. 615, 618 (2004).

“possibility of access” in the case.¹⁹² Unlike *Belton*, where a single officer faced the threat of four unsecured arrestees, *Gant*’s five officers outnumbered its three arrestees, “all of whom had been handcuffed and secured in separate patrol cars” before the time of the search.¹⁹³ The police in *Gant* therefore “could not reasonably have believed . . . that Gant could have accessed his car at the time of the search.”¹⁹⁴

After presenting its “reaching distance” rule, *Gant* turned, with less enthusiasm, to its second innovation: the declaration that “circumstances unique to the vehicle context justify a search incident to a lawful arrest when it is ‘reasonable to believe evidence relevant to the crime of arrest might be found in the vehicle.’”¹⁹⁵ The Court’s lukewarm feelings for the “offense-related evidence” rule¹⁹⁶ were palpable. *Gant* did not bother to explain exactly what circumstances made this rule appropriate for the vehicle context, and it further conceded that the rule “[did] not follow from *Chimel*. ”¹⁹⁷ The Court even tried to minimize the rule’s impact by predicting that “in many cases, as when a recent occupant is arrested for a traffic violation, there will be no reasonable basis to believe the vehicle contains relevant evidence.”¹⁹⁸ The *Gant* Court then proceeded to quickly apply the “offense-related evidence” rule¹⁹⁹ to the facts. Justice Stevens determined that “[a]n evidentiary basis for the search” was lacking because “Gant was arrested for driving with a suspended license—an offense for which police could not expect to find evidence in the passenger compartment of Gant’s car.”²⁰⁰ The search in the case was therefore unreasonable.²⁰¹ Neither of *Gant*’s two new rules—those involving “reaching distance” and “offense-related evidence”—could therefore save the search in the case from violating the Fourth Amendment.²⁰²

192. *Gant*, 129 S. Ct. at 1719.

193. *Id.* *Gant* somehow failed to note the similarities between its arrestees “handcuffed and secured” in patrol cars and the arrestee handcuffed and seated in a patrol vehicle in *Thornton*, a drug case decided only five years previously. *See id.* *Gant* did distinguish *Thornton* on the fact that there was reason to believe that evidence of the offense would be found in the car. *Id.* at 1719, 1722. This second rationale, though, was not recognized by the Court at the time *Thornton* upheld its search incident to arrest, leaving Justice Scalia to advocate such a rule in his concurrence. *See Thornton*, 541 U.S. at 632 (Scalia, J., concurring). Therefore, *Thornton* must have upheld the search based on the belief that police could search vehicles despite having the arrestee safely handcuffed and secured in a patrol car, since this was the only basis at the time to support the search incident to arrest.

194. *Gant*, 129 S. Ct. at 1719.

195. *Id.* (quoting *Thornton*, 541 U.S. at 632 (Scalia, J., concurring)).

196. *Id.*

197. *Id.*

198. *Id.*

199. *Id.*

200. *Id.*

201. *Id.*

202. *Id.* at 1723-24.

III. IMPLICATIONS OF *GANT*

A. Gant's New "Reaching Distance" Limitation on Belton Will Create Uncertainty Both in Its Meaning and Its Impact on Police Behavior

When *Gant* created the rule that an officer may search a vehicle incident to arrest "only if the arrestee is within reaching distance of the passenger compartment at the time of the search,"²⁰³ the Court never precisely defined "reaching distance." Instead, Justice Stevens spent much ink explaining what it was not.²⁰⁴ The facts of the case indicated that "reaching distance" did not extend to an officer meeting the suspect ten to twelve feet from his car, arresting him, handcuffing him, and placing him in the back of a police cruiser.²⁰⁵ The Court also indicated, without elaboration, that "reaching distance" did not occur "in most cases."²⁰⁶ The closest *Gant* came to identifying reaching distance was in dismissing an extreme case: "There was no suggestion . . . that *Chimel* authorizes a vehicle search incident to arrest when there is no realistic possibility that an arrestee could access his vehicle."²⁰⁷

Gant did make some factual determinations about the failure of the case to fulfill reaching distance, but the Court's contentions here created more questions than answers. Justice Stevens distinguished *Belton*, where reaching distance existed, from *Gant*'s own facts (where it did not) by noting that "[u]nlike in *Belton*, which involved a single officer confronted with four unsecured arrestees, the five officers in this case outnumbered the three arrestees, all of whom had been handcuffed and secured in separate patrol cars before the officers searched *Gant*'s car."²⁰⁸ The Court thus deemed relevant three factors: (1) number of officers versus arrestees; (2) handcuffing the arrestee; and (3) placing the arrestee in a patrol vehicle. Yet the Court did not explicitly place the factors in a multi-part test, nor did it explain how many of the factors needed to be satisfied for "reaching distance" to occur.

Interestingly, none of these three factors measures the actual distance an arrestee must be to the vehicle in order to fall within "reaching distance." *Gant*'s discussion of *Belton* lacks any specifics regarding the distance of the arrestees from the vehicle other than a brief mention that the State in *Belton* vaguely referred to the arrestees' "proximity to the vehicle."²⁰⁹ *Gant*—and *Belton*, for

203. *Id.* at 1723.

204. In fact, the first time *Gant* provided its holding, it did so in the negative: "Accordingly, we hold that *Belton* does not authorize a vehicle search incident to a recent occupant's arrest after the arrestee has been secured and cannot access the interior of the vehicle." *Id.* at 1714.

205. *See id.* at 1715.

206. *Id.* at 1719. Specifically, the Court noted that "in most cases the vehicle's passenger compartment will not be within the arrestee's reach at the time of the search." *Id.*

207. *Id.* at 1717.

208. *Id.* at 1719.

209. *Id.* at 1717.

that matter—mentioned the distance of *Belton*'s arrestees in only one other particular: the space between the individuals themselves.²¹⁰ The Court in each case noted that “the officer ‘split . . . [the arrestees] up into four separate areas of the Thruway . . . so they would not be in physical touching area of each other.’”²¹¹ The failure to measure the arrestee’s distance from the vehicle not only undermined this factor’s credibility as an actual element of *Gant*’s “reaching distance” test, but also called into question the importance *Belton* placed on this unspecified variable in the first place. *Gant*’s three-factor rule tests not the distance of the arrestee from the car, but his or her potential danger to the arresting officer in general. Each factor—the number of suspects in relation to the number of officers, the limitation of an arrestee’s movement, whether constrained by handcuffs or locked in a police vehicle—speaks to danger facing police regardless of nearness to a vehicle or the existence of weapons. Instead of “reaching distance,” perhaps *Gant*’s first rule should have been labeled “amount of officer control over arrestee” or “potential danger suspect posed to officer.”

The confusion created by *Gant*’s “reaching distance” rule is more than just semantics. The Court either created a “distance” test without any guidance or a “control/danger” test without acknowledgment. Either option creates uncertainty, which could cause either officer hesitation or exploitation. Hesitation from uncertainty was one of the primary concerns voiced in *Belton*, which recognized that police can only adhere to Fourth Amendment limits if they can understand them.²¹² *Belton* feared that a confusing rule ran the risk of hindering officers in performing their “day-to-day activities.”²¹³ In its vagueness, *Gant*’s “reaching distance” rule might be akin to rules “qualified by all sorts of ifs, ands, and buts and requiring the drawing of subtle nuances and hairline distinctions.”²¹⁴ Officers in the field, groping to figure out the boundaries of “reaching distance,” might be hamstrung in performing their duties.

The uncertainty that might confuse one officer could embolden another to take improper risks. In the wake of *Gant*’s ruling, officers who are intent upon searching a vehicle might attempt to fulfill “reaching distance” by placing the arrestee next to the vehicle. Moreover, *Gant*’s three “reaching distance” factors might create a perverse incentive for police to delay handcuffing an arrestee or placing him in a patrol car. *Gant* may thus result in placing an unrestrained suspect nearer to weapons in a vehicle. Its reasoning, at the very least, promotes proximity between officer and suspect that would trigger the exposure warned about in *Robinson*.²¹⁵ Police and citizen safety might therefore be threatened.

210. See *id.* (citing *New York v. Belton*, 453 U.S. 454, 456 (1981)).

211. *Id.* (quoting *Belton*, 453 U.S. at 456).

212. See *Belton*, 453 U.S. at 458.

213. See *id.*

214. *Id.* (citation omitted).

215. See *United States v. Robinson*, 414 U.S. 218, 234, 234-35 (1973). The *Robinson* Court noted,

It is scarcely open to doubt that the danger to an officer is far greater in the case of the

Finally, the “reaching distance” rule created a curious incongruity between the reasoning officers apply when they are outside a vehicle and the thinking they perform when they are inside a vehicle. As previously noted, *Gant* rejected a bright-line rule enabling officers to search every vehicle after arresting its occupant, instead requiring an officer to assess in each individual case whether the arrestee is near enough to the car to trigger the “reaching distance” requirement to search.²¹⁶ Yet once an arrestee is within “reaching distance” of the vehicle, *Gant* then has the officer shift to *Belton*’s bright-line rule enabling a search of the entire passenger compartment, regardless of the arrestee’s ability to actually reach every area within that passenger compartment.²¹⁷ The *Gant* Court has thus grafted together a case-by-case test with a bright-line rule, reaping the worst of both worlds. Officers will have to toil through all the interpretive factual issues of a case-by-case analysis to determine whether they can even employ the *Belton* rule. Then, should they find they can indeed search the car, police can look in places beyond the arrestee’s reach, thus losing the precision usually offered in a case-by-case rule.

B. Gant’s New “Reasonable to Believe the Vehicle Contains Evidence of the Offense” Standard Expands Police Search Power Beyond Chimel’s Reasoned Basis

In introducing its entirely new basis for police to search a vehicle incident to arrest (the “likelihood of discovering offense-related evidence”), the *Gant* Court somewhat sheepishly noted that its innovation “[did] not follow from *Chimel*.²¹⁸ This was an understatement, for *Gant*’s “offense-related evidence” rule threatened to undo much of *Chimel*’s hard work in anchoring search incident to arrest to “the inherent necessities” of the arrest situation.²¹⁹ The “offense-related evidence” rule stands wholly independent of the “reaching distance” rule and is therefore not subject to its limitations.²²⁰ In his initial pitch for this rule in his *Thornton* concurrence, Justice Scalia made this plain.²²¹ There, he declared, “If *Belton* searches are justifiable, it is not because the arrestee might grab a weapon or evidentiary item from his car, but simply because the car might contain evidence relevant to the crime for which he was arrested.”²²² Instead, the right to search was based on “a more general interest in gathering evidence relevant

extended exposure which follows the taking of a suspect into custody and transporting him to the police station than in the case of the relatively fleeting contact resulting from the typical *Terry*-type stop.

Id.

216. See *Gant*, 129 S. Ct. at 1723.

217. See *id.* at 1719-20.

218. *Id.* at 1719.

219. *Chimel v. California*, 395 U.S. 752, 759 (1969).

220. See *Gant*, 129 S. Ct. at 1719.

221. *Thornton v. United States*, 541 U.S. 615, 629 (2004) (Scalia, J., concurring).

222. *Id.*

to the crime for which the suspect had been arrested.²²³ Thus, when *Gant* adopted Justice Scalia's expansion of search incident to arrest,²²⁴ it presumably did so based on the rationale advanced by Justice Scalia. These reasons were, by Justice Scalia's own admission, a general interest in gathering evidence, not a "more specific interest in preventing its concealment or destruction."²²⁵ Since this rule therefore does not concern itself with the arrestee's reaching distance, it might not only outstrip *Chimel*'s restrictions, but also *Belton*'s. *Belton* crafted the "passenger compartment" limit because "articles inside the relatively narrow compass of the passenger compartment of an automobile are in fact generally, even if not inevitably, within 'the area into which an arrestee might reach in order to grab a weapon or evidentiary item'."²²⁶ Now that *Gant*'s "offense-related evidence" rule is not dependent on the arrestee's reach, police search of the vehicle could presumably move beyond the passenger compartment to include the entire vehicle. Suddenly, motorists would have no basis to complain about police looking into a trunk, under the hood, or behind a door panel. Searches would be limited only by the size of the stopped vehicle, some of which could be large SUVs, motor homes, or tractor-trailers.

Furthermore, *Gant*'s trigger for the offense-related evidence search—"reasonable to believe"²²⁷—is itself problematic. Insertion of the "reasonable to believe" standard into an area of law which has routinely specified the needed level of certainty²²⁸ could create a vacuum, tending to confuse officers, citizens, and judges. Historically, the Court has employed the "reasonable to believe" standard inconsistently—sometimes suggesting probable cause, other times reasonable suspicion, and still other times using the phrase as a shortcut to express the notion that certain crimes might by their nature have evidence of their commission near the arrestee. The *Gant* Court seemed to employ "reasonable to believe" in this later sense; in the passage that the Court quoted from Justice Scalia's concurrence,²²⁹ Justice Scalia offered, "I would . . . limit *Belton* searches to cases where it is reasonable to believe evidence relevant to the crime of arrest might be found in the vehicle."²³⁰ Justice Scalia then

223. *Id.*

224. *Gant*, 129 S. Ct. at 1719, 1721.

225. *Thornton*, 541 U.S. at 629 (Scalia, J., concurring).

226. *New York v. Belton*, 453 U.S. 454, 460 (1981) (alteration in original) (quoting *Chimel v. California*, 395 U.S. 752, 763 (1969)).

227. *Gant*, 129 S. Ct. at 1719.

228. For instance, the Court has specified the need to establish probable cause in a variety of contexts, such as the automobile exception in *Carroll v. United States*, 267 U.S. 132, 149 (1925); search incident to arrest in *Rawlings v. Kentucky*, 448 U.S. 98, 111 (1980); and serving arrest warrants on a home in *Payton v. New York*, 445 U.S. 573, 603 (1980). Likewise, the Court has pinpointed reasonable suspicion as the standard for both *Terry* stops and frisks in *Terry v. Ohio*, 392 U.S. 1, 20-21, 28 (1968); frisks of vehicles in *Michigan v. Long*, 463 U.S. 1032, 1047 (1983); and protective sweeps of homes in *Maryland v. Buie*, 494 U.S. 325, 334 (1990).

229. *Gant*, 129 S. Ct. at 1719.

230. *Thornton*, 541 U.S. at 632 (Scalia, J., concurring). Justice Scalia based his "reasonable

promptly applied his “reasonable to believe” standard to the facts in *Thornton* by noting,

In this case, as in *Belton*, petitioner [Thornton] was lawfully arrested for a drug offense. It was reasonable for . . . [the arresting officer] to believe that further contraband or similar evidence relevant to the crime for which he had been arrested might be found in the vehicle from which he had just alighted and which was still within his vicinity at the time of arrest.²³¹

Justice Scalia’s use of “reasonable to believe” language was meant to focus attention on identifying “the nature of the charge”²³² because certain crimes tend to have evidence of their commission nearby, such as contraband in a drug case or stolen property in a theft case. Thus, Justice Scalia’s “reasonable to believe” did not offer a standard as to level of certainty officers should have before launching into the search.²³³

“Reasonable to believe” has also surfaced in the context of reasonable suspicion.²³⁴ In his dissenting opinion in *Safford v. Redding*, a case involving a strip search of a student, Justice Thomas declared searches of students to be “permissible in scope under *T.L.O.* so long as it is objectively reasonable to believe that the area searched could conceal the contraband.”²³⁵ *New Jersey v. T.L.O.* was itself a school search case which based a school administrator’s

to believe” test in part on use of a similar phrase: “reason to believe.” *Id.* at 630. Earlier in his concurring opinion, Justice Scalia cited Bishop’s *Criminal Procedure*, employing the “reason to believe” language as follows:

The officer who arrests a man on a criminal charge should consider the nature of the charge; and, if he finds about the prisoner’s person, or otherwise in his possession, either goods or moneys which there is *reason to believe* are connected with the supposed crime as its fruits, or as the instruments with which it was committed, or as directly furnishing evidence relating to the transaction, he may take the same, and hold them to be disposed of as the court may direct.

Id. (emphasis added) (citation omitted).

231. *Id.* at 632.

232. *Id.* at 630 (citation omitted).

233. *Robinson* used “reasonable to believe” in a similar fashion, noting that the court of appeals in the case concluded that “there could be no evidence or fruits in the case of an offense such as that with which . . . [Robinson] was charged” since he was arrested for driving on a suspended license. *United States v. Robinson*, 414 U.S. 218, 233 (1973). The court of appeals in *Robinson* would have allowed a search only “[w]here the arrest is made for a crime for which it is reasonable to believe that evidence exists” rather than where the arrest is for driving on a suspended license, where it would not be reasonable to believe that any such evidence would be found. *Id.* at 233 n.4.

234. *Safford Unified Sch. Dist. No. 1 v. Redding*, 129 S. Ct. 2633, 2650 (2009) (Thomas, J., concurring and dissenting).

235. *Id.*

search of a purse upon reasonable suspicion.²³⁶ In his *Safford* opinion, Justice Thomas referred to reasonable suspicion in this context explicitly, noting that the Court had just acknowledged that “school officials had reasonable suspicion to look in Redding’s backpack.”²³⁷ He further declared that such “reasonable suspicion” later “did not dissipate simply because . . . Redding was secreting the pills in a place she thought no one would look.”²³⁸

The Court, however, has more frequently mentioned “reasonable to believe” in the context of probable cause. As early as 1948, the Court mentioned “reasonable to believe”²³⁹ in *Johnson v. United States*, a case in which officers approached a room emitting the “distinctive and unmistakable” odor of burning opium.²⁴⁰ The *Johnson* Court declared that the police, entering what they thought was an “opium smoking den” being used by possibly “one or several persons,” only obtained probable cause to believe that Johnson herself was in possession of opium after they had intruded and found her to be the only person present.²⁴¹ Prior to their entry, officers had no “reasonable basis for believing” Johnson was in possession and thus lacked probable cause for arrest.²⁴² Similarly, in *United States v. Leon*, famous for establishing the good faith exception to the exclusionary rule,²⁴³ the Court spoke of officers properly executing a warrant and searching only “those places and for those objects that it was reasonable to believe were covered by the warrant.”²⁴⁴ Because probable cause is the basis of a search pursuant to a warrant, the “reasonable to believe”²⁴⁵ standard as used in *Leon* apparently rose to the level of certainty required in probable cause. The Court also employed the “reasonable to believe” standard in *Michigan v. Fisher*, a case decided in 2009.²⁴⁶ In *Fisher*, police entered a home after finding “a household in considerable chaos: a pickup truck in the driveway with its front smashed, damaged fenceposts along the side of the property, and three broken house windows, the glass still on the ground outside.”²⁴⁷ Officers noticed blood on the truck and on one of the doors to the home.²⁴⁸ Through a window, police could see Fisher “screaming and throwing things.”²⁴⁹ The Court deemed it “objectively reasonable to believe that Fisher’s projectiles might have a human target (perhaps a spouse or a child), or that Fisher would hurt himself in the

236. *New Jersey v. T.L.O.*, 469 U.S. 325, 341 (1985).

237. *Safford*, 129 S. Ct. at 2650 (Thomas, J., concurring and dissenting).

238. *Id.*

239. *Johnson v. United States*, 333 U.S. 10, 16 (1948).

240. *Id.* at 12.

241. *Id.* at 16.

242. *Id.*

243. *United States v. Leon*, 468 U.S. 897, 922 (1984).

244. *Id.* at 918 n.19.

245. *Id.*

246. *Michigan v. Fisher*, 130 S. Ct. 546, 549 (2009) (per curiam).

247. *Id.* at 547.

248. *Id.*

249. *Id.*

course of his rage” and thus upheld the police entry under “the emergency aid exception.”²⁵⁰ Although the Court refrained from mentioning probable cause, the officers’ direct observations of all relevant details supported such a level of certainty.

The Court’s choice of “reasonable to believe” language is problematic. The phrase has been applied inconsistently in Court precedent, sometimes meaning reasonable suspicion and other times meaning probable cause. Furthermore, “reasonable to believe,” in comparison to the traditional measures of reasonable suspicion and probable cause, has rarely been used by the Court in a Fourth Amendment context. The phrase therefore lacks the clarity that comes with the repeated testing of a rule that is applied to various factual situations. Thus, the “reasonable to believe” language further muddles an already confusing rule.

C. Gant’s New “Reasonable to Believe the Vehicle Contains Evidence of the Offense” Rule Risks Equating Search Incident to Arrest with the Automobile Exception

Even if “reasonable to believe” is interpreted in its narrowest sense as requiring probable cause,²⁵¹ the Court’s new rule, allowing searches for offense-related evidence, still raises troubling questions. If *Gant* intended to enable police to search “the passenger compartment of an arrestee’s vehicle and any containers”²⁵² based on a reasonable belief amounting to probable cause, then the Court essentially grafted a lesser version of the automobile exception onto vehicle searches incident to arrest. The automobile exception, created in *Carroll v. United States*, enables police to search a vehicle without a warrant if probable cause exists to believe it contains “contraband goods in the course of transportation and subject to forfeiture or destruction.”²⁵³

The Court created the automobile exception, a right independent of search incident to arrest, for reasons different from those justifying search incident to arrest.²⁵⁴ The automobile exception was intended to enable officers to cope with

250. *Id.* at 549.

251. Such an interpretation could fit within the holding of *Carroll v. United States*, where the Court declared that

the true rule is that if the search and seizure without a warrant are made upon probable cause, that is, upon a belief, reasonably arising out of circumstances known to the seizing officer, that an automobile or other vehicle contains that which by law is subject to seizure and destruction, the search and seizure are valid.

Carroll v. United States, 267 U.S. 132, 149 (1925).

252. *Arizona v. Gant*, 129 S. Ct. 1710, 1719 (2009).

253. *Carroll*, 267 U.S. at 149. However, the Court later expanded the class of objects to be searched to include instrumentalities and “fruits of the crime.” *Chambers v. Maroney*, 399 U.S. 42, 47 (1970).

254. *Chambers*, 399 U.S. at 49 (explaining, “The Court also noted that the search of an auto on probable cause proceeds on a theory wholly different from that justifying the search incident to an arrest.”).

the exigencies created by the car's mobility as a mode of transportation.²⁵⁵ Later, the Court adopted a second rationale to justify the automobile exception—that cars possess a lessened expectation of privacy due to their use and exposure to official regulation.²⁵⁶ In contrast, the Court, as previously noted, allowed police to search incident to arrest in order to protect their own safety and the evidence of the case.²⁵⁷ Due to different justifications, the two warrant requirement exceptions have differing search rights, both as to time and space.

Although *Chimel* limited the spatial scope of a search to the arrestee's person and area of immediate control,²⁵⁸ and *Belton* interpreted *Chimel* to enable the search of the passenger compartment of a vehicle,²⁵⁹ the Court placed no such limits on the automobile exception. In *Carroll*, the Court simply spoke in terms of searching an automobile.²⁶⁰ Later, in *United States v. Ross*, the Court specified that the probable cause triggering an automobile exception search justified "the search of every part of the vehicle and its contents that may conceal the object of the search," including containers.²⁶¹ The Court has deemed the scope of the automobile exception to be "broad,"²⁶² even allowing officers to tear open the upholstery cushions in a vehicle search.²⁶³ Because *Gant*'s new right to search might be based on probable cause to believe that a vehicle contains offense-related evidence, the Court might expand the scope of search incident to arrest to the broader search permitted under the other search right for vehicles based on

255. See *Carroll*, 267 U.S. at 146 (stating, "It is impossible to get a warrant to stop an automobile. Before a warrant could be secured the automobile would be beyond reach of the officer with its load of illegal liquor disposed of."); see also *Chambers*, 399 U.S. at 50 (stating, "[T]he mobility of a car may make the search of a car without a warrant reasonable 'although the result might be the opposite in a search of a home, a store, or other fixed piece of property'" (quoting *Preston v. United States*, 376 U.S. 364, 366-67 (1964))).

256. In *United States v. Chadwick*, the Court explained that "[o]ne has a lesser expectation of privacy in a motor vehicle because its function is transportation and it seldom serves as one's residence or as the repository of personal effects It travels public thoroughfares where both its occupants and its contents are in plain view." *United States v. Chadwick*, 433 U.S. 1, 12 (1977) (quoting *Cardwell v. Lewis*, 417 U.S. 583, 590 (1974)), abrogated by *California v. Acevedo*, 500 U.S. 565 (1991). *Chadwick* also noted:

Other factors reduce automobile privacy. "All states require vehicles to be registered and operators to be licensed. States and localities have enacted extensive and detailed codes regulating the condition and manner in which motor vehicles may be operated on public streets and highways." Automobiles periodically undergo official inspection, and they are often taken into police custody in the interests of public safety.

Id. at 12-13 (quoting *Cady v. Dombrowski*, 413 U.S. 433, 441 (1973)).

257. *Chimel v. California*, 395 U.S. 752, 763 (1969).

258. *Id.*

259. *New York v. Belton*, 453 U.S. 454, 460 (1981).

260. See *Carroll*, 267 U.S. at 162.

261. *United States v. Ross*, 456 U.S. 798, 825 (1982).

262. *California v. Acevedo*, 500 U.S. 565, 572 (1991).

263. E.g., *Ross*, 456 U.S. at 805.

probable cause—the automobile exception. There would seem to be no logical hurdle to making such a leap. If one search based on probable cause for the existence of evidence allows a search of the entire car, why not the other?

Gant also threatens to disturb the temporal scope of search incident to arrest by confusing this search right with the automobile exception. Just as the spatial boundaries of search incident to arrest and the automobile exception differ, so too do the time limits of these searches. The Court has mandated that a search incident to arrest be “contemporaneous” with the arrest because the justifications for officer safety and preservation of evidence “are absent where a search is remote in time or place from the arrest.”²⁶⁴ The automobile exception is not limited by the same time constraints because officers who have obtained the right to use the automobile exception in the field can choose to exercise it later—that is, when the car is safely secured at the police station.²⁶⁵ The Court has even upheld, on one occasion, the right to perform an automobile exception three days after seizure.²⁶⁶ *Gant*’s new adoption of the offense-related evidence rule could easily expand the temporal scope of search incident to arrest of vehicles to match that of the automobile exception. This is due to the Court’s reliance on unspecified “circumstances unique to the vehicle context”²⁶⁷ to support its new search right. Because the Court never elaborated on what circumstances make vehicles unique, the Court will be free in the future to adopt those identified in the automobile exception context—mobility and lessened privacy expectations. After all, the Court has previously altered the automobile exception itself from a rule based on the exigency of a moving vehicle to one that covers cars immobilized at police stations. Its reasoning has been that the justification to perform an automobile exception search does “not vanish once the car has been immobilized,” for “there is no requirement of exigent circumstances to justify such a warrantless search.”²⁶⁸ A similar evolution of the justification for search incident to arrest would presumably result in similar time boundaries. Thus, *Gant* has conceivably opened the door to expanding search incident to arrest of vehicles to searches days later at police stations.

Perhaps the *Gant* Court, in adding a kind of automobile exception appendage to its search incident to arrest rule, merely assumed it was buying Justice Scalia’s fifth vote on the cheap. After all, the right to search the vehicle would be limited to the passenger compartment, an area already open to officers’ reach by *Belton*. Furthermore, it would apply only in those arrests—such as drug possession or theft—where it would be reasonable to believe evidence of the crime was in the car, information that would trigger the automobile exception anyway. Yet, as shown by the histories of both search incident to arrest and the automobile exception, the limits on these searches hardly remain static. The rationales advanced by *Gant* could make search incident to arrest vulnerable to dramatic

264. Preston v. United States, 376 U.S. 364, 367 (1964).

265. See Chambers v. Maroney, 399 U.S. 42, 52 (1970).

266. See, e.g., United States v. Johns, 469 U.S. 478, 487 (1985).

267. Arizona v. Gant, 129 S. Ct. 1710, 1719 (2009).

268. *Johns*, 469 U.S. at 484 (citations omitted).

expansions already suffered by the automobile exception. The purchase of Justice Scalia's fifth vote might have been made at a steep price.

D. The Logic of Gant's Rule for Searching Vehicles Might Alter All Searches Incident to Arrest

Although *Gant* explicitly limited its "reasonable to believe" rule to "circumstances unique to the vehicle context,"²⁶⁹ the reasoning *Gant* employed could alter the scope of all searches incident to arrest. *Gant*'s impact on the scope of search incident to arrest is particularly troubling when the case is viewed in relation to *Robinson*. *Robinson* involved an officer searching a "crumpled up cigarette package"²⁷⁰ of a driver arrested for operating a motor vehicle after revocation of his operator's permit.²⁷¹ The Court in *Robinson* explicitly rejected the contention that the likelihood of finding a weapon or evidence was relevant to the right to search incident to arrest, asserting,

The authority to search the person incident to a lawful custodial arrest, while based upon the need to disarm and to discover evidence, does not depend on what a court may later decide was the probability in a particular arrest situation that weapons or evidence would in fact be found upon the person of the suspect.²⁷²

In contrast, *Gant* not only considered such probabilities relevant, but also included them in its "reasonable to believe" in the existence of offense-related evidence rule.²⁷³ *Gant*, involving an arrest on essentially the same charges as those in *Robinson*,²⁷⁴ starkly declared that when an officer made an arrest for a traffic offense (like the officers in *Gant* and *Robinson*), officers could not search because "there will be no reasonable basis to believe the vehicle contains relevant evidence."²⁷⁵ *Gant* thus directly contradicted *Robinson*.

Admittedly, *Robinson* and *Gant* addressed distinct search issues; *Robinson* involved the search of the arrestee's person,²⁷⁶ whereas *Gant* focused on the search of the arrestee's vehicle.²⁷⁷ Yet this difference fails to explain the reason for the creation of a bright-line search right in *Robinson* and a factual analysis in *Gant*.²⁷⁸ Did *Gant* mean to retain *Robinson*'s absolute right to search an

269. *Gant*, 129 S. Ct. at 1719.

270. *United States v. Robinson*, 414 U.S. 218, 223 (1973).

271. *Id.* at 220-21.

272. *Id.* at 235.

273. See *Gant*, 129 S. Ct. at 1719.

274. *Id.* at 1714-15 (arresting defendant for driving with a suspended license); *Robinson*, 414 U.S. at 220 (arresting defendant for operating a motor vehicle after revocation of his operating permit).

275. *Gant*, 129 S. Ct. at 1719.

276. See *Robinson*, 414 U.S. at 224.

277. See *Gant*, 129 S. Ct. at 1714-15.

278. See *id.* at 1719 (making the Court's failure to spell out the "circumstances unique to the

arrestee's person—regardless of the possibility of finding weapons or evidence—while at the same time create a “reasonable to believe” rule for searches of passenger compartments of vehicles? *Gant* fails to mention this distinction, let alone justify it. Without a reasoned distinction, *Gant* potentially calls into question police searches of arrestees for evidence if it is not “reasonable to believe evidence relevant to the crime of arrest might be found.”²⁷⁹ The force of *Gant*'s logic could, at worst, undermine officers' abilities to make quick ad hoc judgments as necessary in arrest situations and, at least, sow confusion.

CONCLUSION

Belton warned, “When a person cannot know how a court will apply a settled principle to a recurring factual situation, that person cannot know the scope of his constitutional protection, nor can a policeman know the scope of his authority.”²⁸⁰ *Gant* should have heeded this advice. The *Gant* Court’s attempt to limit search incident to arrest of vehicles to only those cases where an arrestee would have access to weapons or evidence was admirable, yet clumsy. When considering facts deemed relevant in assessing reaching distance, *Gant* offered little guidance in terms of proximity, instead listing factors of control: handcuffing, ratio of officers to arrestees, and placing the arrestee in the police car.²⁸¹ *Gant*’s lack of attention to actual distance might confuse some officers, while its focus on control might induce more reckless officers to take dangerous chances—such as forgoing the use of handcuffs and placing arrestees next to vehicles—in order to justify intrusion into automobiles. Therefore, the practical effect of *Gant* might undermine the very safety rationales supporting the warrant exception in the first place.

If *Gant*’s reaching distance test is unclear, its second rule—enabling officers to search a vehicle based on a reason to believe the automobile contains offense-related evidence—is not only unnecessary, but potentially corrosive to the boundaries of the search incident to arrest rule. Since *Gant* created the offense-related evidence rule as an entirely new basis to search an arrestee’s vehicle,²⁸² there seems to be no reason for believing that this new search right is limited by *Belton*’s passenger compartment boundary. Thus, while the reaching distance rule might limit officers to the passenger compartment of the vehicle, the offense-related evidence rule might not. Further, *Gant* has failed to specify the level of certainty required to establish the reason to believe that offense-related evidence might be in the car. Therefore, the offense-related evidence rule, lacking specific limits, could potentially be used to expand search incident to arrest to the scope of the automobile exception. Finally, *Gant*’s reticence in identifying what

vehicle context” all the more troubling).

279. *Id.* (citation omitted).

280. *New York v. Belton*, 453 U.S. 454, 459-60 (1981).

281. *Gant*, 129 S. Ct. at 1719.

282. *See id.*

“circumstances” were indeed “unique to the vehicle context”²⁸³ could cause future courts to determine that perhaps vehicles are really not “unique,” thus enabling them to apply the offense-related evidence search rule to situations occurring outside the vehicle context.

Justice Scalia, who originally championed a search based on a reason to believe that a vehicle contained offense-related evidence,²⁸⁴ might himself come to regret the ramifications of *Gant*. He declared in the context of a police frisk, a search arguably less intrusive than a search incident to arrest, that he frankly doubted “whether the fiercely proud men who adopted our Fourth Amendment would have allowed themselves” to undergo such an indignity as the Court envisioned.²⁸⁵ Similarly, one wonders whether the Founders, who often doubted the judgment of officials at the scene,²⁸⁶ would welcome an officer searching their carriages based on their own determination that it was “reasonable to believe evidence relevant to the crime of arrest might be found in the vehicle.”²⁸⁷ Perhaps fiercely proud citizens today should be equally concerned about *Gant*, a case of doubtful certainty.

283. *Id.*

284. See *Thornton v. United States*, 541 U.S. 615, 632 (2004) (Scalia, J., concurring).

285. *Minnesota v. Dickerson*, 508 U.S. 366, 381 (1993) (Scalia, J., concurring).

286. *Logan*, *supra* note 23, at 385.

287. *Gant*, 129 S. Ct. at 1719 (quoting *Thornton*, 541 U.S. at 632 (Scalia, J., concurring)).

TOWARDS RECOGNIZING AND RECONCILING THE MULTIPLICITY OF VALUES AND INTERESTS IN TRADEMARK LAW

MICHAEL S. MIRELES, JR.*

TABLE OF CONTENTS

Introduction	428
I. Overview of the Contraction and Expansion of Trademark Rights	435
II. The Policy and Structure of Trademark Law	438
A. <i>The Policy of Trademark Law</i>	439
1. Consumer Search Cost Theory and Consumer Perception	439
a. Consumer search cost theory	439
b. Consumer perception	443
2. The Public Interest and Values	446
B. <i>The Structure of Trademark Law</i>	455
1. Distinctiveness and Secondary Meaning	455
2. Genericness and Genericide	457
3. Strength of the Mark	460
4. Dilution	463
5. Fair Use	467
6. Abandonment	468
7. Equitable Defenses	469
8. Licensing	471
9. Other Reasons to Police	471
10. Conclusion	472
C. <i>The Expansion of Trademark Law and the Inadequacy of the Consumer Search Cost Theory</i>	473
1. Merchandising	473
2. Dilution	476
3. New Theories of Confusion	478
a. Post-sale confusion	478
b. Reverse confusion	480
c. Initial interest confusion	480
d. Sponsorship and association confusion	484
4. Trademark Use	486
5. The <i>Dawn Donut Rule</i>	487
6. Overtaking the Public Domain	488
III. Proposals for Reform	488
A. <i>The Public Interest, Values, and Injunctive Relief</i>	488
1. Preliminary Injunctions	489
2. Permanent Injunctions	497

* Associate Professor of Law, University of the Pacific, McGeorge School of Law. I am grateful for the helpful comments of Robert Brauneis, Gary Pulsinelli, and several reviewers. I am also thankful for the excellent editorial assistance of the *Indiana Law Review* staff. As always, I am responsible for any errors.

B. <i>Eliminating Enforcement Efforts as a Basis for Obtaining, Retaining, and Expanding Rights</i>	499
Conclusion	500

INTRODUCTION

In the last seventy years, trademark rights have expanded enormously.¹ Many commentators believe this has led to an unjustified increase in the rights and remedies available to trademark owners.² This expansion has been approved and led by trademark owners, Congress, and courts, including the U.S. Supreme Court.³ However, in the last ten years or so, the Supreme Court and Congress have begun attempting to restrain trademark rights,⁴ and trademark law is striving to recognize the multiplicity of values implicated by trademark law and reconcile the interests of mark owners and the public generally—particularly users of marks, consumers, and the competitors of mark owners. This Article attempts to explain this search by exploring the sometimes competing interests of mark holders and the public; trademark policy, including the lack of coherent limits inherent in the consumer search cost theory as a normative tool and other important values that are relevant to trademark law; and the structure of trademark law. Following the lead of recent Supreme Court opinions concerning injunctions in *Winter v. Natural Resources Defense Council, Inc.*⁵ and *eBay Inc. v. MercExchange, L.L.C.*,⁶ this Article also provides proposals for addressing the various values and interests in trademark law.

Some light can be shed on the expansion and contraction of trademark law

1. See Glynn S. Lunney, Jr., *Trademark Monopolies*, 48 EMORY L.J. 367, 369-71 (1999) (explaining how trademark protection has expanded since the middle of the twentieth century).

2. See Robert C. Denicola, *Trademarks as Speech: Constitutional Implications of the Emerging Rationales for the Protection of Trade Symbols*, 1982 WIS. L. REV. 158, 159-60; Rochelle Cooper Dreyfuss, *Expressive Genericity: Trademarks as Language in the Pepsi Generation*, 65 NOTRE DAME L. REV. 397, 398 (1990); Alex Kozinski, Essay, *Trademarks Unplugged*, 68 N.Y.U. L. REV. 960, 973 (1993); Mark A. Lemley, *The Modern Lanham Act and the Death of Common Sense*, 108 YALE L.J. 1687, 1688, 1697-713 (1999); Jessica Litman, *Breakfast with Batman: The Public Interest in the Advertising Age*, 108 YALE L.J. 1717, 1722 (1999); Lunney, *supra* note 1, at 367-68.

3. See Daniel M. McClure, *Trademarks and Competition: The Recent History*, 59 LAW & CONTEMP. PROBS. 13, 13 (1996) (“The recent history of trademark law has shown a trend toward greater protection of trademarks and rejection of most claims that trademarks have anticompetitive effects.”).

4. See David S. Welkowitz, *The Supreme Court and Trademark Law in the New Millennium*, 30 WM. MITCHELL L. REV. 1659, 1660 (2004) (arguing that in four cases since 2000, the Supreme Court has signaled its disapproval with “the expansive view of trademark protection put forth by many lower courts” and that this “stem[s] from the Court’s conviction that trademark law remains an offshoot of unfair competition rather than a subset of intellectual property law”).

5. 129 S. Ct. 365, 375-76 (2008).

6. 547 U.S. 388, 394 (2006).

by examining its somewhat elusive purposes and the structure of trademark law itself. First, this Article examines trademark law theory and the inherent limitations and inconsistencies in the development of trademark doctrine that arise from reliance on the consumer search cost theory as a normative device. The Federal Lanham Act broadly directs two purposes of trademark law: first, to protect producers from unfair competition; and second, to protect consumers from deception.⁷ The Lanham Act thus takes into account the interests of mark owners and consumers. The Supreme Court has expanded on these purposes by adopting the consumer search cost theory,⁸ a theory proposed by Judge Richard Posner and Professor William Landes as a positive theory to explain trademark doctrine.⁹ The consumer search cost theory posits that trademark law is designed to lower the costs of consumers in making decisions to purchase goods and services.¹⁰ For example, trademark law lowers search costs for consumers by preventing potential confusion in the marketplace.¹¹ To the extent that one's use of a mark may be confusingly similar to that of a prior user, the senior user may be able to enjoin the use of the junior user. If this confusion were not prohibited by trademark law, consumers might have difficulty locating a particular desired product or service, and thus, their costs in finding the good or service they want would be raised. Therefore, trademarks facilitate consumer choice and foster our market economy.¹² The benefits of trademarks for consumers include the incentive for mark holders to produce goods and services of a consistent quality¹³ and maintain a certain minimum level of quality because trademarks affix

7. Two Pesos, Inc. v. Taco Cabana, Inc., 505 U.S. 763, 767-68 (1992).

8. See Qualitex Co. v. Jacobson Prods. Co., 514 U.S. 159, 163-64 (1995).

9. See William M. Landes & Richard A. Posner, *Trademark Law: An Economic Perspective*, 30 J.L. & ECON. 265, 268-70 (1987) [hereinafter Landes & Posner, *Trademark Law*]. But see Barton Beebe, *The Semiotic Analysis of Trademark Law*, 51 UCLA L. REV. 621, 624-26 (2004) (asserting that economic theory fails to adequately explain trademark doctrine such as distinctiveness and dilution and that semiotic analysis will provide additional insight into trademark law).

10. See WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 167-68 (2003) [hereinafter LANDES & POSNER, ECONOMIC STRUCTURE] (describing the consumer search cost theory).

11. See Beebe, *supra* note 9, at 677 ("The rationale for anti-infringement protection is currently articulated in the terminology of economics and is said to involve the minimization of search costs and the promotion of consistent levels of product quality."); Landes & Posner, *Trademark Law*, *supra* note 9, at 300-06 (explaining infringement and confusion theories).

12. See Margreth Barrett, *Trademarks and Digital Technologies: "Use" on the Net*, 13 NO. 11 J. INTERNET L. 1, 1 (2010) [hereinafter Barrett, *Digital Technologies*] (noting that "[t]rademark rights promote an efficient, competitive marketplace").

13. See LANDES & POSNER, ECONOMIC STRUCTURE, *supra* note 10, at 167 ("[T]he benefits of trademarks in reducing the cost to consumers of distinguishing among brands of a product require that the producer of a trademarked good maintain a consistent quality of his output, that is, that he make sure that from the consumer's standpoint it really is the same product from unit to unit and time to time.").

responsibility to a single source. However, a focus on the reduction of consumer costs requires an assessment of consumer perception and fails to take into account other values and interests implicated by trademark law—and therein lies the problem.¹⁴ Because mark holders can control consumer perception through advertising and enforcement, they are able to determine the extent of protection their marks may possess.¹⁵ Thus, the consumer search cost theory may not provide adequate limits to trademark law, and relying upon it will continue to facilitate its expansion. Indeed, Professor Mark McKenna has persuasively argued that reframing trademark law as a means to reduce consumer search costs has failed to stop the expansion of new trademark rights.¹⁶

Moreover, there are public interests and values implicated by trademark law that may not be adequately protected by courts. For example, the public has an interest in ensuring that consumers have access to information about goods and services from competing producers. This information includes that which truthfully compares and describes the goods and services of both parties.¹⁷ Thus, trademark law should ensure that competitors have access to information they need to facilitate competition by providing that information to consumers.¹⁸ Another important public interest and value implicated by trademark law is the protection of free speech. This interest strongly resonates with the interest of users of trademarks and other members of the public.¹⁹ Users of trademarks

14. See Mark P. McKenna, *Trademark Use and the Problem of Source*, 2009 U. ILL. L. REV. 773, 821 [hereinafter McKenna, *Trademark Use*] (“[M]odern trademark law, deeply influenced by the search costs rationale, regards consumer confusion as a problem in and of itself. Not surprisingly, consumer understanding therefore pervades trademark law.”).

15. See Mark P. McKenna, *The Normative Foundations of Trademark Law*, 82 NOTRE DAME L. REV. 1839, 1916 (2007) [hereinafter McKenna, *Normative Foundations*] (“Consumer expectations now carry all the weight for those who hope to limit trademark protection, and those expectations have proven almost infinitely pliable. Producers are able to frame just about any argument for broader protection in terms of consumer expectations, which they are in position to influence systematically through marketing. Moreover, once courts and Congress began to expand trademark law and committed it to consumer understanding, expansion became self-reinforcing—broader protection begets consumer expectations of great control, which begets even more protection.”).

16. See *id.*; McKenna, *Trademark Use*, *supra* note 14, at 821.

17. See Barrett, *Digital Technologies*, *supra* note 12, at 1 (“Rights in marks should not prevent competing producers from effectively communicating the nature, qualities, and characteristics of their own products to interested consumers or prevent competitors, consumers, or the media from engaging in critical product critiques and commentary.”).

18. The search cost theory can explain trademark doctrine to the extent that certain defenses, such as the nominative fair use defense or the explicit comparative advertising exclusion for dilution, allow some use of another’s trademark in comparative advertising. See 15 U.S.C. § 1125(c)(3) (2006); Century 21 Real Estate Corp. v. LendingTree, Inc., 425 F.3d 211, 214 (3d Cir. 2005).

19. The interests described as the user interest and competitor interest can also be explained to some extent under the search cost theory. For example, competitors need to allow for the

include those who do not offer competing goods but wish to comment on a particular brand with criticism or use the mark to make some other socially relevant point about another topic.²⁰ Users of trademarks may also have an interest in preventing the protection of a particular trademark by one entity because it may be scandalous or disparaging.²¹ The public also has an interest in ensuring that trademark law does not unnecessarily raise barriers to entry of new firms and that trademark law is not abused to the detriment of the public by extending trademark protection to items that should be in the public domain.²² A rigid application of the search cost theory as an underlying justification for trademark law may underprotect and overprotect the interests of the public—including the consumer, user, and competitor—and the mark owner. This Article reviews these purposes and goals from the perspective of the particular benefited stakeholder.²³

This Article asserts that courts should expressly unpack and account for the interests and values that are implicated by trademark disputes.²⁴ Failure by

provision of additional information that reduces the consumer's costs in making purchasing decisions. The user interest is less straightforward but could be described in the following way under a broad vision of the consumer search cost theory: through protection of free speech by users, the consumer acquires critical information about a producer that may impact the consumer's decision to purchase a product or service. For example, a consumer may discover that a corporation sells products intended for children that are manufactured by companies that employ people who work incredibly long hours, thus impacting workers' ability to make quality products. With that information, the consumer may wish to avoid the risk that a corporation does not adequately ensure that the products it sells are not potentially dangerous and purchase the products he wants from another retailer.

20. See, e.g., Laura A. Heymann, *The Public's Domain in Trademark Law: A First Amendment Theory of the Consumer*, 43 GA. L. REV. 651, 656-58 (2009) (advocating a view of the consumer grounded in autonomy theory, recognizing that “the consumer [should] be left free to make whatever associations she wants with the marks she encounters, even if those associations are not the ones the mark holder would prefer, or not the ones that would be optimal from the perspective of the individual’s intellectual or personal development,” particularly as against theories of trademark law “that depend in part on the persuasive value of the mark rather than on its source-identifying aspects”).

21. See generally *Harjo v. Pro-Football, Inc.*, 50 U.S.P.Q.2d (BNA) 1705 (T.T.A.B. 1999).

22. See generally *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23 (2003); *Fleischer Studios, Inc. v. A.V.E.L.A., Inc.*, No. 09-56317, 2011 WL 631449 (9th Cir. Feb. 23, 2011); Shubha Ghosh, *Dilution and Competition Norms: The Case of Federal Trademark Dilution Claims Against Direct Competitors*, 24 SANTA CLARA COMPUTER & HIGH TECH. L.J. 571 (2008).

23. For example, the protection of consumers from deception benefits the public, and the protection of goodwill benefits the producer, although there is overlap between the purposes. See *infra* notes 76-90 and accompanying text.

24. This notion has been expressed by other commentators who point to the use of judicially crafted defenses to curb excesses in trademark protection; however, at least one commentator has expressed skepticism concerning the realistic development of defenses by a judiciary that is increasingly formalistic. Compare Graeme B. Dinwoodie, *Developing Defenses in Trademark Law*,

courts to account for the interests and values they are balancing and who the interest benefits can result in an expansion of trademark protection based on an unstated (or sometimes stated) desire to protect a mark from free riding without an adequate assessment of other interests that may need consideration.²⁵ The question arises as to where and when courts should assess those interests and values.²⁶ In light of recent Supreme Court cases *Winter*²⁷ and *eBay*,²⁸ this Article suggests that analysis should occur at the preliminary and permanent injunction stage.²⁹

Second, this Article analyzes the structure of the law and how it encourages producers to enforce their marks vigorously against third party uses of their marks so that courts will declare that a mark is valid and/or entitled to a broader scope of protection.³⁰ For example, the strength of the mark in the likelihood of confusion analysis examines evidence of third party usage of a mark.³¹ This encourages mark holders to enforce their marks against third parties to shut down that use. The requirement of fame for dilution also pushes mark holders to reduce third party usage of their marks.³² For example, although a mark may not

13 LEWIS & CLARK L. REV. 99 (2009) (arguing for the development of defenses that take into account values other than those driven purely by consumer confusion concerns), *with* Michael Grynberg, *Things Are Worse Than We Think: Trademark Defenses in a “Formalist” Age*, 24 BERKELEY TECH. L.J. 897, 901-02 (2009) (arguing that formalist courts are unlikely to expand or develop defenses to address the concerns raised by Professors Dinwoodie and Janis).

25. Cf. Robert G. Bone, *Hunting Goodwill: A History of the Concept of Goodwill in Trademark Law*, 86 B.U.L.REV. 547, 622 (2006) [hereinafter Bone, *Goodwill*] (“Sometimes broad liability may be justified because of the high enforcement costs of tailoring protection to fit information transmission policies more closely. But broad liability should never be justified simply on the ground that trademark law prevents goodwill appropriation. Until this form of justification is eliminated, we cannot hope to achieve a sensible and coherent body of trademark law.”).

26. Professor William McGeveran states that “almost all recent controversial cases protected speech,” but the real problem is the chilling of speech by threats of legal action; thus, “the priority should be restructuring the relevant doctrines to reduce the pre-litigation chilling effect.” William McGeveran, *Four Free Speech Goals for Trademark Law*, 18 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1205, 1206-07 (2008) [hereinafter McGeveran, *Free Speech*]; see also Pierre N. Leval, *Trademark: Champion of Free Speech*, 27 COLUM. J.L. & ARTS 187 (2004) (arguing that trademark law doctrine effectively protects free speech interests).

27. *Winter v. Natural Res. Def. Council, Inc.*, 129 S. Ct. 365 (2008).

28. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388 (2006).

29. This Article does not argue that trademark law should not continue to account for interests such as free speech in the application of the likelihood of confusion test or any defenses. *See Facenda v. N.F.L. Films, Inc.*, 542 F.3d 1007, 1015-18 (3d Cir. 2008) (considering First Amendment defense); *Lamparello v. Falwell*, 420 F.3d 309, 313-15 (4th Cir. 2005) (discussing defenses to dilution actions and likelihood of confusion test that protect free speech interests). The public interest should still be addressed under those doctrines.

30. See *infra* Part II.B.

31. See *infra* notes 177-86 and accompanying text.

32. See *infra* notes 187-205 and accompanying text.

yet be famous under the Trademark Dilution Revision Act, mark holders³³ are provided an incentive to reduce third party usage by filing infringement actions or sending cease and desist letters to make it more likely that their marks will become famous. In order to avoid genericism, mark holders must also police third party usage of their marks to reduce the chance that a court will find a mark generic.³⁴ In fact, courts have found that mere evidence of policing is probative of whether a mark has become generic.³⁵ Finally, in determining if a mark will receive protection as a descriptive mark with secondary meaning or if a mark is descriptive or suggestive, courts will examine whether competitors are using the mark.³⁶ This may provide a further incentive for mark holders to shut down competitors using a possibly confusingly similar mark.

With the availability of trademark rights, the scope of those rights, and the potential loss of those rights in the partial control of the trademark holder, it is not surprising that trademark owners will push for maintaining their rights and strengthening existing rights.³⁷ Indeed, we want trademark owners to enforce their rights vigorously because, as discussed below, this should protect the public from confusion and enhance the ability of consumers to confidently make choices about what goods and services they wish to obtain. However, if courts are unable to fully appreciate and incorporate the public's interests, then courts will be inclined to continue to expand rights—particularly when more confusion is likely as mark holders change consumers' perceptions about the existence and scope of their rights. This expansion is particularly pernicious in areas involving values such as free speech, which, if not adequately taken into account by courts, may lead to a chilling effect in future cases by risk-averse users.³⁸ Adding to the

33. In referring to "mark holders" or "mark owners," I am generally including the risk-averse lawyers who counsel these mark holders or owners.

34. See *infra* notes 167-76 and accompanying text.

35. See *infra* notes 167-76 and accompanying text.

36. See *infra* notes 155-66 and accompanying text.

37. Cf. McKenna, *Trademark Use*, *supra* note 14, at 821-22 ("[T]his heavy focus on consumer understanding renders trademark law inherently unstable. Consumers' expectations naturally evolve as they become more familiar with new commercial contexts or relationships. . . . [L]egal doctrine and consumer expectations feed off each other, creating an endless loop: what consumers know (or think they know) about the law shapes expectations, which then feed back into the law only to influence future expectations.").

38. Lisa P. Ramsey, *Increasing First Amendment Scrutiny of Trademark Law*, 61 SMU L. REV. 381, 384-85 (2008) ("Protected expression is frequently suppressed or chilled by trademark law because the law's current built-in First Amendment safeguards, such as the descriptive fair use doctrine, are limited and involve fact-specific determinations that often can only be resolved after discovery at summary judgment or trial—a cost many defendants cannot afford."); see also McGeveran, *Free Speech*, *supra* note 26, at 1206-07 ("Considerable anecdotal evidence suggests that the real action occurs outside the courthouse: markholders send cease-and-desist letters and threaten legal action against those using trademarks to facilitate speech, and the recipients frequently capitulate."). For a discussion of abusive trademark litigation, see K.J. Greene, *Abusive Trademark Litigation and the Incredible Shrinking Confusion Doctrine—Trademark Abuse in the*

problem is that most issues concerning public perception are intensely factual, and thus, they are generally not amenable to early resolution through summary judgment.³⁹ It therefore becomes critically important for courts to explain their reasoning for granting relief against a purported trademark infringer or diluter. Importantly, in an empirical study, Professor Kenneth Port found that the number of reported trademark cases in recent years has decreased substantially while the number of filed trademark cases has increased substantially.⁴⁰ Professor Port argues that the difference could be based on trademark holders forcing alleged infringers and diluters to settle early—trademark extortion.⁴¹ Notably, the United States Patent and Trademark Office has requested “feedback from U.S. trademark owners, practitioners, and others regarding their experiences with litigation tactics, especially those involving an attempt to enforce trademark rights beyond a reasonable interpretation of the scope of the rights granted to the trademark owner.”⁴² The danger of the chilling of activity beneficial to the public is great.

This Article proposes that courts consider the interests of the trademark owner and public, along with other values, at the preliminary injunction stage. This procedure will allow trademarks suits to be disposed of early enough in litigation to reduce defendants’ costs and incentivize the early resolution of suits. This Article also suggests some additional proposals that may tip the balance toward the public over the mark holder. However, these proposals are conservative because we should not tilt too far against the mark holder due to the mark holder’s important role in enforcing the public interest in reducing consumer deception in the market and in business development.

First, courts must define the relative interests of the public and producer and

Context of Entertainment Media and Cyberspace, 27 HARV. J.L. & PUB. POL’Y 609 (2004).

39. See *Clicks Billiards Inc. v. Sixshooters Inc.*, 251 F.3d 1252, 1267 (9th Cir. 2001) (“This case underscores our warning that ‘trial courts disfavor deciding trademark cases in summary judgments because the ultimate issue is so inherently factual.’” (quoting *Levi Strauss & Co. v. Blue Bell, Inc.*, 778 F.2d 1352, 1355 n.5 (9th Cir. 1985))); see also *Dinwoodie*, *supra* note 24, at 99 (arguing for the development of defenses that take into account values other than those driven purely by consumer deception concerns); cf. Robert G. Bone, *Enforcement Costs and Trademark Puzzles*, 90 VA. L. REV. 2099, 2134-36 (2004) [hereinafter Bone, *Enforcement Costs*] (explaining the rule that presumes confusion where there is a direct competitor and the same mark based on a reduction of “administrative costs and eliminat[ion of] . . . erroneous acquittals and their associated costs”).

40. Professor Port collected reported trademark cases from 1947 to 2005 and noted that “[t]rademark litigation has seen a precipitous drop since 2001.” Kenneth L. Port, *Trademark Extortion: The End of Trademark Law*, 65 WASH. & LEE L. REV. 585, 622 (2008).

41. *Id.* at 632. Another explanation is that infringers and diluters could be forcing trademark owners to settle early.

42. See *Request for Comments: Trademark Litigation Tactics*, U.S. PATENT & TRADEMARK OFFICE, http://www.uspto.gov/trademarks/bullies_survey.jsp (last visited Mar. 22, 2011). The Secretary of Commerce was directed to investigate trademark litigation tactics by Congress. See Trademark Technical and Conforming Amendment Act of 2010, Pub. L. No. 111-146, 124 Stat. 66.

the relevant values in applying the public interest prong when determining whether to issue preliminary and permanent injunctions. The preliminary injunction stage is critical in trademark litigation because it occurs early in the suit, and the denial or grant of the injunction will push parties to settle. If a court is required to consider the public interest and values and balance those interests against the mark owner's interest, the court may deny a preliminary injunction because the public's interest—not necessarily the potential infringer's or diluter's interests—outweighs the mark holder's interest. This outcome may lead to settlement, perhaps using methods such as disclaimers to mitigate consumer confusion.⁴³ There is still a problem with this proposal if many potential infringers/diluters decide to capitulate and change their actions based on a cease and desist letter. However, after a number of decisions are reported that account for the public interest and values relevant to trademark law and deny relief, a potential infringer/diluter may litigate through the preliminary injunction stage and before costly discovery. This Article also argues that determinations concerning permanent injunctive relief should engage in similar balancing.

The next proposal also attempts to address some of the incentives that mark holders have to bring trademark suits in the first instance. In assessments concerning the scope and existence of rights, such as secondary meaning, genericism, strength of the mark, and fame, courts should not consider evidence of policing efforts by mark holders.⁴⁴ This change may provide some dampening effect on mark holders' enforcement efforts. At the very least, mark holders will not have as much of a direct incentive under trademark law itself to send cease and desist letters and file infringement suits. However, because of the nature of expansion of trademark rights based on consumer perception, trademark holders will likely continue to enforce their marks. As discussed below, not all enforcement is unwanted because trademark holders vindicate the public's interest through litigation and in some ways may act as private attorneys general. Indeed, we allocate the costs of litigating trademark matters to the trademark holders—although, as discussed below, it is in their best interest to litigate.

First, this Article provides an overview of the contraction and continued expansion of trademark rights. Second, this Article addresses how the policy and structure of trademark law has resulted in the development of trademark doctrine that is overinclusive. Finally, this Article provides proposals to address the overprotection of trademarks.

I. OVERVIEW OF THE CONTRACTION AND EXPANSION OF TRADEMARK RIGHTS

The ebb and flow of the scope of trademark protection has quickened as courts, Congress, and trademark holders react to new technology, the pressures of globalization, and the increasing importance of the brand in national and international markets.⁴⁵ After Congress enacted the Federal Trademark Dilution

43. See *infra* note 321 and accompanying text.

44. See *infra* notes 335-42 and accompanying text.

45. See, e.g., Jerre B. Swann, Sr. et al., *Trademarks and Marketing*, 91 TRADEMARK REP.

Act of 1995,⁴⁶ the Supreme Court restricted the availability of that antidilution law.⁴⁷ Congress, in turn, clarified and attempted to address ambiguities within the Federal Trademark Dilution Act of 1995 by passing the Trademark Dilution Revision Act of 2006 (TDRA).⁴⁸ In part, the TDRA cabined the rights of trademark holders by limiting antidilution protection to a small group of powerful marks and adding specific defenses to dilution.⁴⁹ However, at the same time, the TDRA provided that actual dilution was not required for injunctive relief or damages and that a tarnishment-based action was explicitly available under federal law.⁵⁰ Similarly, with trade dress, lower courts (with the apparent approval of the Supreme Court) began to apply trade dress protection expansively, extending its coverage to designs that could inhibit competition rather than promote it.⁵¹ The Supreme Court reacted by restricting the availability of trade dress protection for product design to limited circumstances.⁵² The Supreme Court further attempted to limit the scope of trade dress protection by clarifying the functionality doctrine and the impact of utility patents on the functionality analysis.⁵³ When some lower courts restrictively applied the classical fair use defense, the Supreme Court responded by expanding the scope of the defense.⁵⁴ Mark owners responded to inadequacies in the protection of their marks under then-current law by stretching trademark doctrine

787, 807 (2001) (“[S]trong brands . . . (i) allow access to consumers’ minds; (ii) make advertising less expensive or more impactful (or both); (iii) enable a manufacturer to communicate more directly with a consumer, cushioning any vagaries of distribution; (iv) assist in attaining channel power; (v) provide a more efficient and credible means of extending into related goods, and give rise to licensing opportunities; (vi) serve as certificates of ‘authenticity’; (vii) afford resilience; and (viii) constitute an asset—brand equity—that is frequently a company’s most valuable single property.”).

46. Federal Trademark Dilution Act of 1995, Pub. L. No. 104-98, 109 Stat. 985 (codified as amended at 15 U.S.C. § 1125(c) (2006)). Although several states have enacted dilution laws, this Article is primarily concerned with trademark rights under federal law.

47. See *Moseley v. V Secret Catalogue, Inc.*, 537 U.S. 418, 434 (2003) (requiring a demonstration of actual dilution and casting doubt on the existence of the tarnishment cause of action under the Federal Trademark Dilution Act).

48. Trademark Dilution Revision Act of 2006, Pub. L. No. 109-312, 120 Stat. 1730 (codified as amended at 15 U.S.C. § 1125(c)).

49. 15 U.S.C. § 1126(c).

50. *Id.*

51. *Two Pesos, Inc. v. Taco Cabana, Inc.*, 505 U.S. 763, 775 (1992).

52. *Wal-Mart Stores, Inc. v. Samara Bros., Inc.*, 529 U.S. 205, 215-16 (2000); see Bone, *Enforcement Costs*, *supra* note 39, at 2155-82 (criticizing the protection of product design trade dress, even with secondary meaning.).

53. *TraFFix Devices, Inc. v. Mktg. Displays, Inc.*, 532 U.S. 23, 30-33 (2001). Courts are still experiencing difficulty in applying the functionality doctrine. See Amy B. Cohen, *Following the Direction of TraFFix: Trade Dress Law and Functionality Revisited*, 50 IDEA 593 (2010) (noting circuit splits on functionality doctrine post-*TraFFix*).

54. *KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc.*, 543 U.S. 111, 123 (2004).

to address cybersquatting,⁵⁵ Congress subsequently responded with an attempt to provide specific relief for mark holders in the cybersquatting context.⁵⁶

For other arguable expansions in the rights of mark holders, the Supreme Court and Congress have not fully responded. And these expansions have marched on. In addition to developing new theories such as post-sale, initial interest, and reverse confusion,⁵⁷ courts have expanded confusion-based theories of liability to related goods and services instead of restricting liability to competing goods and services.⁵⁸ Mark owners can obtain incontestability status,⁵⁹ limiting the assertion of certain defenses against a mark.⁶⁰ Mark owners can reserve a mark through an intent-to-use application, arguably limiting the importance of the concept of “use” in American trademark law as a way to allocate ownership of trademark rights.⁶¹ Mark owners can obtain constructive nationwide rights through federal registration on the Principal Register.⁶² Mark

55. See, e.g., *Panavision Int'l, L.P. v. Toeppen*, 141 F.3d 1316, 1327 (9th Cir. 1998).

56. Anticybersquatting Consumer Protection Act, Pub. L. No. 106-113, 113 Stat. 1501 (1999) (codified as amended in scattered sections of 15 U.S.C.).

57. See *infra* notes 246-72 and accompanying text.

58. *Yale Elec. Corp. v. Robertson*, 26 F.2d 972, 974 (2d Cir. 1928); *Aunt Jemima Mills Co. v. Rigney & Co.*, 247 F. 407, 410-12 (2d Cir. 1917); see also Mark P. McKenna, *Testing Modern Trademark Law's Theory of Harm*, 95 IOWA L. REV. 63, 76-79, 97-115 (2009) [hereinafter McKenna, *Theory of Harm*] (discussing conceptual shift from recognizing infringement between competing goods and services to non-competing goods and services and marketing literature concerning brand extension). Congress arguably affirmed this expansion by modifying the federal Lanham Act in 1962. See *infra* notes 273-77.

59. 15 U.S.C. §§ 1065, 1115(b) (2006).

60. If a mark has been on the Principal Register for at least five years, it may obtain incontestability status upon application by the mark holder. *Id.* § 1065. Once a mark is incontestable, many defenses to infringement may be foreclosed. See *Park 'N Fly, Inc. v. Dollar Park & Fly, Inc.*, 469 U.S. 189, 197-201 (1985). Indeed, a mark holder can sue for infringement and use incontestability status in an offensive way by keeping a potential infringer from raising the defense that a mark is descriptive and never should have been registered. See *id.* at 197-98. Incontestability status is viewed as “quieting title” to a mark, but it instead may harm competition by not allowing other mark holders to use descriptive marks or at least deterring them from using those marks. Incontestability status may overprotect marks beyond what is needed under a consumer search cost theory, and it may require a basis under an anti-free riding principle; however, that may not be as defensible as a basis for trademark protection under the consumer search cost theory. See *infra* notes 82-90, 119-20 and accompanying text. For two critiques of incontestability, see generally Suman Naresh, *Incontestability and Rights in Descriptive Trademarks*, 53 U. CHI. L. REV. 953 (1986) and Kenneth L. Port, *The Illegitimacy of Trademark Incontestability*, 26 IND. L. REV. 519 (1993).

61. 15 U.S.C. § 1051(b). For additional discussion, see Stephen L. Carter, Comment, *The Trouble with Trademark*, 99 YALE L.J. 759, 781-85 (1990).

62. 15 U.S.C. §§ 1072, 1115(a). For additional discussion, see Carter, *supra* note 61, at 788-95.

holder rights have also expanded because of the erosion of the *Dawn Donut*⁶³ rule addressing the availability of injunctive relief between geographically remote users where consumer confusion is unlikely.⁶⁴ Mark owners have benefited from the dismantling of the “use in commerce” requirement for trademark infringement.⁶⁵ Trademark holders can license trademarks with very limited evidence of actual quality control.⁶⁶ Trademark owners can also obtain so-called “merchandising rights” and perhaps even drive the scope of their own protection through enforcement actions and by directing consumer perception through advertising.⁶⁷ In sum, the subject matter of trademark protection has continued to expand.⁶⁸ And the expansion of the protection for trademarks is likely to continue unabated.

II. THE POLICY AND STRUCTURE OF TRADEMARK LAW

This section reviews the policy underlying trademark law and explains how

63. *Dawn Donut Co. v. Hart's Food Stores, Inc.*, 267 F.2d 358, 360 (2d Cir. 1959) (holding that a mark holder was not entitled to relief when the other party was in a “separate trading area[]” and no “present likelihood . . . [of expanding into owner’s] market” area existed).

64. See Robert C. Cumbow, *Use Is the New Protectability, Dawn Donuts Are Still Hot This Season, and Other Trademark Issues*, 1.4 LANDSLIDE 20 (2009).

65. See *Rescuecom Corp. v. Google Inc.*, 562 F.3d 123, 127-31 (2d Cir. 2009); Stacey L. Dogan, *Beyond Trademark Use*, 8 J. ON TELECOMM. & HIGH TECH. L. 135, 136 (2010) (noting that post-*Rescuecom*, “[i]n the Second Circuit, at least, the trademark use requirement for infringement is all but dead”); Stacey L. Dogan & Mark A. Lemley, *Trademarks and Consumer Search Costs on the Internet*, 41 Hous. L. Rev. 777, 780 (2004) [hereinafter Dogan & Lemley, *Search Costs*] (noting recent decisions where courts have expanded the concept of trademark “use” in the infringement context “to include . . . [alleged infringers who] did not even arguably offer their own products or services under the mark”).

66. See *infra* notes 221-23 and accompanying text.

67. Cf. Stacey L. Dogan & Mark A. Lemley, *The Merchandising Right: Fragile Theory or Fait Accompli?*, 54 EMORY L.J. 461, 485-86 (2005) [hereinafter Dogan & Lemley, *Merchandising Right*] (“It is possible that consumers have come to expect that San Francisco Giants jerseys are licensed by the Giants, not because they serve a brand-identifying function but simply because the law has sometimes required such a relationship.”); Lemley, *supra* note 2, at 1708 (noting the same effect).

68. This expansion has occurred in part because of the emphasis of trademark law on the consumer search cost theory. Because a consumer—any human—can use almost anything as a symbol or device to indicate source, the expansion of trademark law to include various types of trade dress such as product design, color, sounds, and smells is not surprising. Indeed, recently, color alone has received protection upon a showing of secondary meaning. *See Qualitex Co. v. Jacobson Prods. Co.*, 514 U.S. 159, 174-75 (1995) (holding that a trademark holder has the ability to create trademark rights in a color by manipulating the perceptions of consumers through the creation of acquired distinctiveness). For a criticism of the expansion of trademark protection to include color, see generally Ann Bartow, *The True Colors of Trademark Law: Greenlighting a Red Tide of Anti Competition Blues*, 97 KY. L.J. 263 (2009).

it fails to provide adequate limitations on the rights available to mark holders. It further analyzes how the structure of trademark law itself, given the focus of trademark law on consumer perception, results in the vigorous enforcement and policing of trademark rights by trademark holders. These two problems have enabled the expansion of trademark rights against a background of new technology such as the Internet, globalization, and the increased importance and value of the brand.

A. The Policy of Trademark Law

This section generally discusses the consumer search cost theory and points out some problems with relying upon consumer perception to create trademark doctrine. This section also discusses some of the interests and values that courts should expressly consider in deciding whether to issue a preliminary or permanent injunction.

1. Consumer Search Cost Theory and Consumer Perception.—

a. Consumer search cost theory.—The Lanham Act⁶⁹ is intended to serve two purposes with respect to trademark law: the protection of consumers from deception and the protection of mark holders from unfair competition.⁷⁰ The dominant view of the purpose of trademark law is the law and economics approach: the reduction of consumer search costs.⁷¹ The Supreme Court has recently stressed the importance of trademarks in reducing consumer search costs,⁷² which can lead to an incentive to produce products of consistent quality.⁷³

69. Pub. L. No. 79-489, 60 Stat. 427 (codified as amended in sections of 15 U.S.C.).

70. Two Pesos, Inc. v. Taco Cabana, Inc., 505 U.S. 763, 767-68 (1992).

71. See Landes & Posner, *Trademark Law*, *supra* note 9, at 268-70.

72. *Qualitex Co.*, 514 U.S. at 163-64; see also *Union Nat'l Bank, Laredo v. Union Nat'l Bank, Austin*, 909 F.2d 839, 844 (5th Cir. 1990) (referencing a purpose of trademark law as reducing consumer costs and deterring free riders); LANDES & POSNER, *ECONOMIC STRUCTURE*, *supra* note 10, at 168 (“The value of a trademark to the firm that uses it to designate its brand is the saving in consumers’ search costs made possible by the information that the trademark conveys or embodies about the quality of the firm’s brand. The brand’s reputation for quality and thus the trademark’s value depend on the firm’s expenditures on quality, service, advertising, and so on.”). Interestingly, most courts do not refer to the reduction of consumer search costs as a primary goal but focus on protecting consumers from deception and ensuring that mark holders realize the benefit from the investment in their goodwill. Both of those goals are generally consistent with the consumer search cost theory.

73. *Qualitex Co.*, 514 U.S. at 163-64. The reduction of consumer search costs can lead to the development of quality products. *Id.* at 164. However, it is more accurate to state that it leads to the development of “consistent quality.” See LANDES & POSNER, *ECONOMIC STRUCTURE*, *supra* note 10, at 168 (“When a brand’s quality is inconsistent, consumers learn that the trademark does not enable them to relate their past to their future consumption experiences; the trademark does not reduce their search costs; they are unwilling to pay more for the branded than for the unbranded good; and so the firm will not earn a sufficient return on its expenditures on promoting the trademark to justify making them.”).

The Supreme Court has also expressed concerns that trademark law is not used in anticompetitive ways, such as inhibiting the ability of new entrants to the market to compete effectively.⁷⁴ While the consumer search cost theory is attractive because it explains much of trademark doctrine, it fails to provide meaningful limits to trademark law because it is based on consumer perception—which is largely under the control of the mark owner.⁷⁵

The consumer search cost theory is apparently primarily concerned with the consumer, although it also protects the trademark owner's interest to a limited extent. This theory posits that trademark law should serve to reduce the costs of consumers in making purchasing decisions by preventing deception and confusion⁷⁶ or (some might argue) by preventing the whittling away of the distinctiveness or tarnishment of a mark.⁷⁷ Legal protection for trademarks thus fosters our market economy by allowing consumers to confidently (and very quickly) select or avoid the goods and services that they have previously enjoyed or disliked.⁷⁸ This provides an incentive to trademark owners to produce goods

Some marks may purposefully signify a lower quality (and perhaps less expensive) product, and others may signify a higher quality (although more expensive) product. So, importantly for the consumer, what trademark law ably supplies is an indication that goods are of the same quality—that he will get what he had before. Trademarks also affect quality by serving to affix responsibility, for example, for a defective good. This provides an incentive to provide a certain level of quality.

74. Two Supreme Court cases—*Two Pesos* and *Wal-Mart Stores v. Samara Bros.*—provide examples of the Court's concern with new entrants. In *Two Pesos*, the Court expressed concern with the ability of new entrants to protect the development of goodwill through the availability of protection for inherently distinctive trade dress. *Two Pesos*, 505 U.S. at 775. In *Wal-Mart Stores*, the court pointed to the ability of established mark holders to use anticompetitive “strike suits” to threaten new entrants to the market with trade dress actions based on inherently distinctive product design. *Wal-Mart Stores, Inc. v. Samara Bros., Inc.*, 529 U.S. 205, 213-14 (2000). In *Two Pesos*, the Court appeared to sanction the protection of product design and packaging trade dress as potentially inherently distinctive. See *Two Pesos*, 505 U.S. at 770. In *Wal-Mart Stores*, however, the court restricted trade dress protection for product design with established acquired distinctiveness. *Wal-Mart Stores*, 529 U.S. at 212. For a discussion of competition norms in trademark law, see generally Ghosh, *supra* note 22, at 572.

75. McKenna, *Normative Foundations*, *supra* note 15, at 1916.

76. See LANDES & POSNER, ECONOMIC STRUCTURE, *supra* note 10, at 167 (“To perform its naming function a trademark or brand name . . . must not be duplicated.”).

77. Some do not believe that the consumer search cost theory provides an adequate basis for dilution. See *infra* notes 237-45 and accompanying text. However, for simplicity at this point, I assume that consumer search cost theory does support some legal protection for blurring-type dilution. For a discussion of blurring-type dilution, see *infra* Part II.B.4.

78. See Landes & Posner, *Trademark Law*, *supra* note 9, at 270 (“The value of a trademark is the saving in search costs made possible by the information or reputation that the trademark conveys or embodies about the brand (or the firm that produces the brand). . . . Once the reputation is created, the firm will obtain greater profits because repeat purchasers and word-of-mouth references will generate higher sales and because consumers will be willing to pay higher prices for

of a consistent quality.⁷⁹ Thus, the consumer search cost theory appears to be primarily concerned with the protection of consumers first, and the protection of goodwill is subordinated to that concern. As a result, we rely on the producer to protect its goodwill when threatened by bringing trademark suits and thus protect the public through the reduction of consumer confusion and dilution. Accordingly, to some extent, the interests of the producer and the consumer are aligned.⁸⁰ The producer will enforce its trademark rights if there is a likelihood of confusion and dilution and, thus, harm to goodwill.⁸¹ However, if there is no clearly actionable consumer confusion or dilution, then the consumer search cost theory may not provide a very effective rationale for the legal protection of trademarks in all circumstances. In that case, an additional theoretical basis must be asserted to support the legal protection of trademarks beyond confusion or dilution—such as preventing free riding or ensuring that there is an adequate incentive to invest in the creation of marks themselves.⁸² That theory must

lower search costs and greater assurance of consistent quality."); *see also* LANDES & POSNER, ECONOMIC STRUCTURE, *supra* note 10, at 167 ("Rather than reading the fine print on the package to determine whether the description matches his understanding of brand *X*, or investigating the attributes of all the different versions of the product (of which *X* is one brand) to determine which one is brand *X*, the consumer will find it much less costly to search by identifying the relevant trademark and purchasing the corresponding brand.").

79. *See* LANDES & POSNER, ECONOMIC STRUCTURE, *supra* note 10, at 167 ("[T]he benefits of trademarks in reducing the cost to consumers of distinguishing among brands of a product require that the producer of a trademarked good maintain a consistent quality of his output, that is, that he make sure that from the consumer's standpoint it really is the same product from unit to unit and time to time.").

80. *See* Levi Strauss & Co. v. Genesco, Inc., 742 F.2d 1401, 1403 (Fed. Cir. 1984) ("Like all good laws, those relating to trademarks are primarily focused on the public interest. The law looks to the pocketbook interests of trademark owners, rather than to a government agency, as the genesis of enforcement efforts, and to the common interest of the trademark owner and the public in the prevention of purchaser confusion."); Glynn S. Lunney, Jr., *Trademarks and the Internet: The United States' Experience*, 97 TRADEMARK REP. 931, 932 (2007) ("[W]ith classic source confusion, where company *A* uses company *B*'s trademarks to sell its products in a manner that leads ordinarily prudent consumers to buy the products of company *A* believing they were made by company *B*, an injunction both protects consumers from confusion and vindicates the trademark owner's legitimate interests in its marks.").

81. *See* Vincent Chiappetta, *Trademarks: More Than Meets the Eye*, 2003 U. ILL. J.L. TECH. & POL'Y 35, 43 (noting that individual consumers are "an unsatisfactory vehicle for policing the integrity of marketplace signals . . . [because of the] minimal interest of any individual and the decentralized nature of the harm" and that we therefore give "the first seller/user of a mark the right to prevent confusingly similar subsequent adoptions. The resulting 'ownership' permits trademark law to harness the seller/user's self-interest in avoiding competition as the mechanism for ensuring enthusiastic attention to threats to the market information system."). Another choice would be complete reliance on the government to enforce marks.

82. At least three commentators have attempted to justify the protection of trademarks under an investment and/or protection against free riding rationale. *See id.* at 37-38 (proposing an

recognize and support some value in the mark or the goodwill behind the mark—*independent from the search cost theory*—and that value must outweigh the potential harms to competition from over-enforcing trademarks by preventing price competition and the entry of competitors.⁸³

From the producer perspective, a theory divorced from consumer search cost that provides theoretical support for broader trademark protection is attractive, given the importance and value of the brand in a global marketplace.⁸⁴ With market values of brands in the billions, it is not surprising that producers vigorously protect and defend their marks.⁸⁵ Indeed, some of the expansion in trademark doctrine can be attributed to an aggressive assertion of trademark rights by producers.⁸⁶ Additionally, there is a historical basis for enforcing trademarks rights primarily as a means to prevent the diversion of sales, thus protecting producer interests—not protecting the consumer interest.⁸⁷ However,

investment/encouragement rationale for trademarks in the context of the modern brand); David J. Franklyn, *Debunking Dilution Doctrine: Toward a Coherent Theory of the Anti-Free-Rider Principle in American Trademark Law*, 56 HASTINGS L.J. 117, 117 (2004) (arguing for a dilution action based on preventing free riding); Simone A. Rose, *Will Atlas Shrug? Dilution Protection for “Famous” Trademarks: Anti-Competitive “Monopoly” or Earned “Property” Right?*, 47 FLA. L. REV. 653, 663 (1995) (pointing out that antidilution rights are based on an earned property right theory). *But see* Wendy J. Gordon, *On Owning Information: Intellectual Property and the Restitutionary Impulse*, 78 VA. L. REV. 149, 157 (1992) (“[T]he potential free riders—the users, copyists, and adapters—are not mere parasites. Many are creators themselves. They may reach markets different than those reached by the original creators, or they may bring new perspective, reduced cost, special expertise, deeper insights, or innovative technology to the exploitation and adaptation of established works.”); William P. Kratzke, *Normative Economic Analysis of Trademark Law*, 21 MEMPHIS ST. U. L. REV. 199, 220-21 (1991) (“Promoting hard work for its own sake and deterring ‘free-riding’ for its own sake eventually injure interbrand competition. An inviolable property right to use a particular trademark in whatever manner eventually impedes interbrand competition by denying others the opportunity to promote their competitive interests while not promoting the competitive interest of the property holder. The competitive interest of all competitors is to benefit consumers.”).

83. This Article argues that the balancing can take place under the public interest prong of the test for the issuance of a preliminary and permanent injunction. Professor Bone states that some trademark expansion can be explained by examining the costs of enforcing trademark law, “including the administrative costs of adjudicating trademark lawsuits and the error costs of over- and under-enforcing trademark rights.” Bone, *supra* note 39, at 2101 (specifically examining product design trade dress and infringement liability expanded to include related products and services).

84. See Chiappetta, *supra* note 81, at 38.

85. See *Best Global Brands*, INTERBRAND, <http://www.interbrand.com/en/best-global-brands/Best-Global-Brands-2010.aspx> (last visited Mar. 24, 2011) (explaining that some brands are valued in the billions of dollars).

86. See *infra* Part II.C.

87. See McKenna, *Normative Foundations*, *supra* note 15, at 1841 (arguing that historically trademark law was not concerned with protecting consumer interests, but “protect[ing] producers

Professor Robert Bone has examined the concept of goodwill in depth and rejects a “general rule, principle, or presumption prohibiting free riding on goodwill.”⁸⁸ Moreover, as discussed below, because of the structure of trademark law and its focus on consumer perception, it may be unnecessary to provide trademark owners with more protection because they already control the scope of their protection under a consumer search cost theory.⁸⁹ Thus, to provide meaningful limits, courts developing trademark doctrine and enforcing trademark rights need to expressly take into account other values and interests against those of the producer to avoid the implicit (and sometimes explicit) sole reliance on prohibiting free riding.⁹⁰

b. *Consumer perception.*—Consumer search cost theory is primarily based on a consideration of consumer perception, and thus, the development of trademark doctrine has followed those considerations.⁹¹ For example, a mark can include anything that a consumer may attach significance or meaning to, which can include any letter, name, product design, or even a color or scent.⁹² There are

from illegitimate diversions of their trade by competitors”). *But see* Bone, *Goodwill*, *supra* note 25, at 560-61 (“In nineteenth century trademark law . . . the two goals [of protecting sellers from loss due to deceptive practices and protecting the public from deception] were mutually consistent and reinforcing—trademark law gave remedies to sellers and in so doing helped both sellers and consumers.”).

88. *See* Bone, *Goodwill*, *supra* note 25, at 621 (“What is problematic is the adverse effect on the mark’s capacity to communicate information to the market. One must balance this cost against the benefit of allowing the use, and the legal rights that will result will necessarily be more limited than those misappropriation alone would justify.”). The definition of goodwill is elusive and depends on the context. *See* LOUIS ALTMAN & MALLA POLLACK, CALLMANN ON UNFAIR COMPETITION, TRADEMARKS AND MONOPOLIES § 1:11 (4th ed. 2009).

89. *See* Chad J. Doellinger, *A New Theory of Trademarks*, 111 PENN. ST. L. REV. 823, 823-24 (2007) (criticizing the consumer search cost theory as a normative theory for trademark law and proposing that trademark law focus on ensuring fair business competition). Although Professor Landes and Judge Posner used search cost theory to explain trademark law, commentators view the reduction of search costs as a normative goal of trademark law. *See* Landes & Posner, *Trademark Law*, *supra* note 9, at 268-70; *cf.* Dogan & Lemley, *Search Costs*, *supra* note 65, at 838 (“Courts must be mindful . . . of the fundamental normative goal that underlies . . . [longstanding trademark rules]—the reduction of consumer search costs.”).

90. *See* Dinwoodie, *supra* note 24, at 99 (arguing for the development of defenses that take into account values other than those driven purely by consumer confusion concerns). The question becomes how one can take into account those interests and whether one can do so early enough to facilitate the early resolution of suits.

91. Some commentators have noted this phenomenon. *See supra* note 67 and accompanying text; Kratzke, *supra* note 82, at 209 (“The search cost of a consumer who sees the same trademark used across different product lines by different sellers may or may not increase depending upon the consumer’s perceptions concerning such use. These perceptions affect the role that trademarks might play in a market.”); McKenna, *Normative Foundations*, *supra* note 15, at 1916.

92. 15 U.S.C. § 1127 (2006) provides:

The term “trademark” includes any word, name, symbol, or device, or any combination

limiting rules as to what can serve as a trademark, but these rules are also often based upon how a consumer perceives the mark.⁹³

The requirement of distinctiveness for trademarks is rooted in the idea that there are some symbols that consumers will realize are source-identifying and some that they will realize are not.⁹⁴ For example, consumers will understand that descriptive words, color, or product design may not necessarily be source-identifiers, but they may instead be there for some aesthetic or descriptive purpose.⁹⁵ The law recognizes, however, that over time, with a significant investment of resources, a producer can change the way consumers perceive descriptive words, color, and product design, leading them to attach a “secondary

thereof—

- (1) used by a person, or
- (2) which a person has a bona fide intention to use in commerce and applies to register on the principal register established by this chapter, to identify and distinguish his or her goods, including a unique product, from those manufactured or sold by others and to indicate the source of the goods, even if that source is unknown.

Id. That section also provides a similar definition for service marks. *Id.*

93. Professor Dinwoodie ably described the role of trademark law and consumer perception, stating, “[S]hould trademark law be structured *reactively* to protect whatever consumer understandings or producer goodwill develops, or should it *proactively* seek to shape the ways in which consumers shop and producers sell or seek to acquire rights, thus shaping how the economy functions?” Graeme B. Dinwoodie, *Trademarks and Territory: Detaching Trademark Law from the Nation-State*, 41 Hous. L. Rev. 885, 889-90 (2004) (emphases added).

94. See Zatarains, Inc. v. Oak Grove Smokehouse, Inc., 698 F.2d 786, 790-92 (5th Cir. 1983), *abrogated by* KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc., 543 U.S. 111 (2004); *see also* Abercrombie & Fitch Co. v. Hunting World, Inc., 537 F.2d 4, 9 (2d Cir. 1976). In a recent study challenging the current legal approach to distinctiveness, Professors Lee, DeRosia, and Christensen, using theoretical tools from the consumer psychology field, hypothesize that “consumers will rely primarily on non-linguistic visual cues (e.g., the location and size of the mark as displayed on a product package) rather than the mark’s semantic meaning.” See Thomas R. Lee et al., *An Empirical and Consumer Psychology Analysis of Trademark Distinctiveness*, 41 ARIZ. ST. L.J. 1033, 1038 (2009). The commentators found, in a series of empirical studies, that “descriptive marks are no less source-indicating than suggestive, arbitrary, and fanciful marks . . . [and] that it is the non-lexical cues of typical trademark use (e.g., the location and size of the mark) that cause consumers to find descriptive marks to be source-indicating.” *Id.* at 1038 (internal citation omitted). Based on their research, the commentators propose the abandonment of “the *Abercrombie* dichotomy between ‘inherently distinctive’ and ‘non-inherently distinctive’ word marks, eliminating (at the threshold protectability stage) the longstanding categories of descriptive, suggestive, arbitrary, and fanciful trademarks . . . [and] all non-generic word marks would be eligible for protection so long as they satisfy the ‘trademark use’ requirement.” *Id.* (emphasis added).

95. See Wal-Mart Stores, Inc. v. Samara Bros., Inc., 529 U.S. 205, 212 (2000); *Abercrombie & Fitch Co.*, 537 F.2d at 9-10.

meaning” to a particular descriptive word, color, or product design.⁹⁶ Conversely, genericide expressly recognizes that over time, the understanding of consumers about a particular mark may change, and the mark may no longer be protected.⁹⁷ Furthermore, the touchstone test for infringement—the likelihood of confusion test—attempts to measure whether there is likely to be confusion amongst an appreciable number of consumers.⁹⁸ Infringement is troubling because it includes not just confusion as to source, but confusion as to sponsorship or affiliation—perhaps broader theories of confusion,⁹⁹ which are also judged from the perspective of the consumer.¹⁰⁰ The scandalousness and disparagement provisions are also examined from the consumer’s or relevant group’s view.¹⁰¹

Nominative fair use and classical fair use also incorporate elements of consumer perception. In nominative fair use, a court may assess whether there is sponsorship or association confusion from the perspective of the consumer.¹⁰² The presence or likelihood of consumer confusion may inform whether a particular use is in good faith under the classical fair use inquiry.¹⁰³ Also, the possible protection of geographic terms is sometimes based upon whether consumers would think there is a goods/place association with a mark.¹⁰⁴ In dilution, courts may inquire whether consumers believe a potential diluting use is likely to cause tarnishment or blurring, and evidence of actual tarnishment or blurring is relevant in evaluating the likelihood of its existence.¹⁰⁵ Finally, the deceptiveness of a mark is also determined from the perspective of the consumer, including whether the purported deception is material to the consumer.¹⁰⁶

In connection with the notion that we rely on trademark holders to uphold the public’s interest through private litigation, it becomes clearer that trademark

96. *Zatarains*, 698 F.2d at 791.

97. See *Murphy Door Bed Co. v. Interior Sleep Sys., Inc.*, 874 F.2d 95, 101 (2d Cir. 1989).

98. See *Daddy’s Junky Music Stores, Inc. v. Big Daddy’s Family Music Ctr.*, 109 F.3d 275, 280 (6th Cir. 1997).

99. For a critical discussion of confusion based on sponsorship or affiliation as a basis for trademark infringement, see Mark A. Lemley & Mark McKenna, *Irrelevant Confusion*, 62 STAN. L. REV. 413 (2010).

100. See *infra* notes 274-76 and accompanying text.

101. See generally *Harjo v. Pro-Football, Inc.*, 50 U.S.P.Q.2d (BNA) 1705 (T.T.A.B. 1999).

102. See *Toyota Motor Sales, U.S.A., Inc. v. Tabari*, 610 F.3d 1171, 1176 (9th Cir. 2010); *Century 21 Real Estate Corp. v. LendingTree, Inc.*, 425 F.3d 211, 218 (3d Cir. 2005); *New Kids on the Block v. News Am. Publ’g, Inc.*, 971 F.2d 302, 308 (9th Cir. 1992).

103. *KP Permanent Make-Up, Inc. v. Lasting Impressions I, Inc.*, 543 U.S. 111, 123 (2004) (“It suffices to realize that our holding that fair use can occur along with some degree of confusion does not foreclose the relevance of the extent of any likely consumer confusion in assessing whether a defendant’s use is objectively fair.”).

104. *In re Nantucket, Inc.*, 677 F.2d 95, 99-101 (C.C.P.A. 1982).

105. See, e.g., 15 U.S.C. § 1125(c)(2)(B)(vi) (2006).

106. See *In re Budge Mfg. Co.*, 857 F.2d 773, 775 (Fed. Cir. 1988).

holders tend to initiate actions to the extent consistent with their desire to protect their marks. These enforcement actions may influence how a consumer perceives marks and thus affect the scope of protection of the marks.¹⁰⁷ Indeed, mark holders have an incredible amount of control over consumer perception by advertising; they make the choice (at least initially) of what images and information are associated or not associated with marks.¹⁰⁸ The creation of a label announcing a licensing relationship between producers can expand rights once consumers understand that entities must obtain permission before using another's mark.¹⁰⁹ The concept of secondary meaning is based on the ability of a mark holder to change consumer perception about the meaning of a particular mark. The circumstantial evidence that courts often rely on includes advertising expenditures using the mark¹¹⁰—the amount of advertising by a particular producer is used to infer whether consumer perception has actually changed.¹¹¹ Thus, consumer search cost theory, which is based in part on an assessment of consumer perception, has resulted to some extent in the ability of producers to manipulate the scope of their trademark rights through the expenditure of more resources—and for that reason, it fails to inherently provide meaningful limits to those rights.¹¹²

2. *The Public Interest and Values.*—For the most part, courts utilize one of the two intended purposes of the Lanham Act—and sometimes, the consumer search cost theory—in analyzing issues involving trademarks. This analysis is incomplete because it fails to expressly account for all of the interests of stakeholders, including users, consumers, and competitors who are involved in many trademark suits.¹¹³ For example, the public has several interests that

107. Cf. James Gibson, *Risk Aversion and Rights Accretion in Intellectual Property Law*, 116 YALE L.J. 882, 907 (2007). Professor Gibson notes that trademark licensing practices create a feedback loop leading to expansive trademark protection. *Id.* He explains that, in part, this is because of the ambiguity of the case law concerning “sponsorship” or “approval.” *Id.*

108. See McKenna, *Normative Foundations*, *supra* note 15, at 1916; see generally Barton Beebe, *Search and Persuasion in Trademark Law*, 103 MICH. L. REV. 2020 (2005).

109. Cf. Dogan & Lemley, *Merchandising Right*, *supra* note 67, at 485 (“[I]f consumers do in fact value obtaining goods from the trademark owner itself—perhaps because it supports the school or team—then we would expect the market to reflect that by developing a distinction between ordinary merchandise and officially licensed merchandise.”).

110. Zatarains, Inc. v. Oak Grove Smokehouse, Inc., 698 F.2d 786, 795 (5th Cir. 1983), *abrogated by* KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc., 543 U.S. 111 (2004).

111. The Lanham Act allows mark owners to use the ® symbol if the mark is registered on the Supplemental Register and does not have secondary meaning. See 15 U.S.C. §§ 1094, 1111 (2006). It is doubtful that consumers would understand this point, however, when they see the ® symbol. Consumers likely believe that the federal government has granted some enforceable “rights” in that particular mark.

112. See Kratzke, *supra* note 82, at 209; McKenna, *Normative Foundations*, *supra* note 15, at 1916; McKenna, *Trademark Use*, *supra* note 14, at 773.

113. Cf. McClure, *supra* note 3, at 32 (“The appeal of the Chicago School economic theory is that it has the capacity to provide an all-encompassing and unifying approach to virtually every

converge with preventing consumer confusion but may need to be divorced from consumer search cost theory to fully appreciate the importance of the particular interest.¹¹⁴ The following portion of this section explains what additional interests may not be completely protected by the consumer search cost theory and provides some examples of how various doctrines do not adequately protect or harm those interests.

First, this section will discuss the ability of competitors to use words to fairly describe their goods and services. Second, it will analyze the right of competitors to use marks in comparative advertising. Finally, this section will discuss other important public interests such as free expression.

The ability of competitors to use a mark to fairly describe their own products and services is an important public interest. Although this public interest is closely tied to the consumer search cost theory, there are some reasons why it should be considered a separate public interest. For example, the goal of lowering consumer search costs can explain why the law protects certain symbols as marks that are distinctive and does not protect marks that consumers do not perceive as distinctive.¹¹⁵ If a non-distinctive mark were protected (e.g., a descriptive mark without acquired distinctiveness), then the mark owner could prohibit a competitor from using a descriptive term that a competitor might need to fairly describe its product. Competitors might need to use that word because it facilitates consumers' ability to find and select goods. Consumers should not be prohibited from using that word because purported mark holders are trying to obtain some non-reputation-related competitive advantage over their competitors.¹¹⁶ Thus, by not allowing protection for descriptive marks without acquired distinctiveness, the consumer search cost theory fits well with the public interest. Allowing such protection would raise consumer search costs by taking

legal issue in trademark law. The function of trademark law is reduced to a single goal of economic efficiency to maximize wealth. . . . This economic approach places less emphasis on other ideas that have historically been viewed as animating goals of trademark law, such as commercial morality, preventing consumer deception, and protecting a trademark owner's business goodwill from misappropriation." (internal citation omitted)). Some of these interests, such as the ability to engage in comparative advertising, are grounded in competition-related concerns beyond reducing consumer search costs.

114. See Deborah R. Gerhardt, *Consumer Investment in Trademark*, 88 N.C. L. REV. 427, 430-31 (2010) ("Courts continue to state that they are honoring the two traditional trademark policies of protecting consumers and mark owners. However, the actual alignment often breaks apart, especially when consumers seek to use marks as information tools. The expansion of trademark law is resulting in trends that ignore or harm consumer interests." (internal citations omitted)).

115. See LANDES & POSNER, ECONOMIC STRUCTURE, *supra* note 10, at 188-97 (discussing how the consumer search cost theory explains the classification of marks on the spectrum of distinctiveness).

116. See *id.* at 189 ("If one producer is allowed to appropriate the word that describes a key attribute, he will obtain rents measured by the higher price he receives for his branded product because he will have made it more costly for his rivals to inform their customers of the attributes of their brands without using the same descriptive word.").

away words from consumers who need the words to make decisions about purchasing goods and services. However, what is troubling is that a descriptive word can be dominated by a particular trademark owner without changing the primary significance of the word or symbol. For example, after five years of substantially continuous and exclusive use, a mark is presumed to have acquired distinctiveness and may be federally registered.¹¹⁷ This presumption, coupled with evidence of sales and advertising, may be sufficient for a mark owner to obtain protection for a mark throughout the United States. This may make it too easy to obtain mark protection, as it removes the mark from use by competitors. Similarly, a mark that has been registered on the Supplemental Register for five years may be presumed to have acquired distinctiveness.¹¹⁸ Indeed, a mark holder who registers a mark on the Principal Register that was wrongly determined by an examiner to have secondary meaning may prevent a competitor from challenging that mark as merely descriptive after five years. Also, the mark holder may use incontestability status offensively in litigation.¹¹⁹ These presumptions may reduce the enforcement costs of trademarks or enable the mark owner to “quiet title,” but they also shift the ability to use particular marks towards the mark owner.

Moreover, these rules are unlikely to be defensible when considering enforcement costs. For example, there is a risk of false positives in finding protectable marks in the case of foreclosing a challenge to a mark based on descriptiveness. The risk of false positives outweighs the benefits of the judicial resources saved by not examining the presence of acquired distinctiveness

117. 15 U.S.C. § 1052(f) (2006). A mark registered on the Principal Register obtains the advantage of one of two presumptions: the mark is not merely descriptive or generic; or, if the mark is descriptive, it has acquired secondary meaning. *See Packman v. Chi. Trib. Co.*, 267 F.3d 628, 638-39 (7th Cir. 2001). Notably, section 1212.05(b) of the *Trademark Manual of Examining Procedure* provides that “[t]he five years of use does not have to be exclusive, but may be ‘substantially’ exclusive. This makes allowance for use by others that may be inconsequential or infringing, which does not necessarily invalidate the applicant’s claim.” U.S. PATENT & TRADEMARK OFFICE, TRADEMARK MANUAL OF EXAMINING PROCEDURE § 1212.05(b) (6th ed. 2009) (citing *L.D. Kichler Co. v. Davoil, Inc.*, 192 F.3d 1349, 1352 (Fed. Cir. 1999)). One court requires “significantly probative evidence” of invalidity to defeat the presumption of validity. *See Borinquen Biscuit Corp. v. M.V. Trading Corp.*, 443 F.3d 112, 118 (1st Cir. 2006).

118. 15 U.S.C. § 1052(f).

119. *See Park ’N Fly, Inc. v. Dollar Park & Fly, Inc.*, 469 U.S. 189, 203 (1985) (determining that incontestability status can be used to foreclose the defense that a mark is descriptive without secondary meaning and should not have been registered in the first place). Incontestability status is obtained through a filing of a § 15 affidavit at the United States Patent and Trademark Office along with five years of use. 15 U.S.C. § 1065. Incontestability status can also be used to determine mark strength under the likelihood of confusion analysis. *Dieter v. B & H Indus. of Sw. Fla., Inc.*, 880 F.2d 322, 328-29 (11th Cir. 1989). An alleged infringer may still argue that it is using the mark in a way protected by the classical fair use defense or the nominative fair use defense. However, the classical fair use and nominative fair use defenses are narrow and vague. *See Bone, supra* note 39, at 2130-34.

because courts would likely examine similar evidence in determining the likelihood of confusion.¹²⁰

Also, while a protected mark may still be used in a descriptive way in good faith to describe a particular good or service under the classical fair use defense, the threat of suit may deter a particular competitor from using that mark in connection with similar goods or services.¹²¹ Moreover, good faith may not exist where a competitor of the mark holder knows of the mark's existence, thus making the defense possibly inapplicable.¹²² Finally, the presence or absence of confusion is still relevant to an analysis of whether an allegedly infringing mark is used in good faith.¹²³

The next public interest is the ability of competitors to use a mark for comparative purposes. Consumer search cost theory may also explain why a competitor may need to use another's mark to make a particular point about its own goods or services. This is because the consumer search cost theory is, in part, concerned with maximizing the amount of productive information provided to consumers. In some contexts, this is best accomplished by allowing the competitor to use the mark owner's mark. Thus, if an entity wants to use another's mark to compare its products or services with a competitor's products or services, it may do so under the nominative fair use theory for infringement or the comparative advertising exclusion for dilution. The use also may be allowed if a court finds no likelihood of confusion. Comparative information is valuable to a consumer in deciding which product or service has the qualities or characteristics that he may desire.¹²⁴ However, it should be noted that nominative fair use has not been adopted in all jurisdictions, and one of the first iterations of

120. Cf. Bone, *supra* note 39, at 2130-34 (explaining how presumptions of secondary meaning for inherently distinctive marks is justified when enforcement costs are analyzed).

121. See Port, *supra* note 40, at 631; see also William McGeeveran, *Rethinking Trademark Fair Use*, 94 IOWA L. REV. 49, 61-63 (2008) [hereafter McGeeveran, *Fair Use*] ("[I]t is difficult to ascertain precisely which doctrines will determine if a particular expressive use qualifies as fair use; the boundaries between them are not clear, and different jurisdictions recognize different forms and combinations of the rules. . . . [T]he various trademark fair use doctrines do not facilitate early adjudication in practice. . . . In this uncertain legal environment, rational markholders get aggressive." (internal citations omitted)).

122. See *infra* notes 206-07 and accompanying text.

123. There has been criticism of the Supreme Court's decision in *KP Permanent Make-Up, Inc.* because it continues to allow the likelihood of confusion to be relevant in a determination of classical fair use. See, e.g., Graeme W. Austin, *Tolerating Confusion About Confusion: Trademark Policies and Fair Use*, 50 ARIZ. L. REV. 157, 185 (2008) ("[I]nsisting on the continued relevance of likelihood of confusion may limit opportunities for summary adjudication where the fair use defense is raised, a result likely to favor proprietors of trademarks that are susceptible to others' descriptive uses.").

124. Comparative advertising in keyword cases has not been allowed by courts in some circumstances. See Lunney, *supra* note 80, at 964-71 (discussing conflicting cases in the context of purchasing keywords for search engines).

the test has been criticized recently by a federal appellate court.¹²⁵ Also, the scope of the law allowing comparative advertising to occur—for example, the nominative fair use approach to infringement—may be too narrow and vague. This may also be true of the comparative advertising defense to a dilution cause of action.¹²⁶ For example, for nominative fair use to apply, confusion as to sponsorship or endorsement is still taken into account, which again implicates the mark holder's ability to control consumer perception.¹²⁷ Even assuming that this approach exists, mark owners will likely attempt to enforce their marks—particularly against small companies—even if those marks are used for comparative advertising purposes. Given the value of the brand to the mark holder, it is plausible that a small company could receive a cease and desist letter and stop the challenged use of the mark even for comparative advertising. The same concerns exist with using the likelihood of confusion analysis to protect the use of a mark for comparative advertising. This impact is particularly troubling where an entity is asserting a mark that perhaps should not receive protection in the first place.¹²⁸

The public interest also includes protecting uses of trademarks that may implicate such values as free expression and ensuring that trademarks are not used to raise prices above those set by a competitive market. This protection may be at odds with the mark holder's interest in protecting the goodwill in its mark. The exact definition of goodwill is somewhat elusive, but current valuations of some marks are in the billions of dollars.¹²⁹ It is not surprising that producers want to ensure that their marks are, and continue to be, protected. Thus, producers protecting their marks may assert actions against users of trademarks whose uses should be protected by free speech values.¹³⁰ Moreover, a producer

125. Circuits that have adopted the nominative fair use approach have disagreed as to its scope. See McGeveran, *Fair Use*, *supra* note 121, at 61-62; see also 21 Century Real Estate Corp. v. LendingTree, Inc., 425 F.3d 211, 228-31 (3d Cir. 2005) (criticizing the Ninth Circuit's version of the test to determine nominative fair use and offering its own version as an affirmative defense); Health Grades, Inc. v. Robert Wood Johnson Univ. Hosp., Inc., 634 F. Supp. 2d 1226, 1239-43 (D. Colo. 2009) (discussing the treatment of the nominative fair use "defense" amongst circuits and disagreeing with the approach taken by the Ninth and Third Circuits).

126. Courts could also find that comparative advertising with a trademark owner's mark can be allowed because there may not be a likelihood of confusion. This depends on how the alleged infringer's mark and the trademark owner's mark are actually used.

127. See *New Kids on the Block* v. News Am. Publ'g, Inc., 971 F.2d 302, 308 (9th Cir. 1992) (explaining that for nominative fair use to apply, "the user must do nothing that would, in conjunction with the mark, suggest sponsorship or endorsement by the trademark holder").

128. See *supra* notes 120-25 and accompanying text.

129. See *Best Global Brands*, *supra* note 85 (stating that the Coca-Cola brand is worth over sixty-eight billion dollars and Google is worth over thirty-one billion dollars).

130. For an example of a trademark case where free speech interests may be implicated, see *Planned Parenthood Fed'n of Am., Inc. v. Bucci*, No. 97-Civ-0629, 1997 WL 133313, at *12 (S.D.N.Y. Mar. 24, 1997), *aff'd*, 152 F.3d 920 (2d Cir. 1998) (unpublished table disposition) (issuing a preliminary injunction enjoining the use of www.plannedparenthood.com and "Planned

may be able to use its mark to raise prices above what a competitive market would set for that particular good or service because of the strength of the mark.¹³¹ Therefore, consumers would be denied the benefit of lower prices for products because of the continued existence of a very strong mark sustained by advertising and enforcement.¹³² Additionally, trademark owners may assert trademark rights over subject matter ordinarily protected by copyright and patent law that is in the public domain, threatening the current balance struck by Congress in enacting those laws.¹³³ There is thus a strong public interest in protecting the public domain and ensuring that the public receives the benefits of competition.

For the most part, the responsibility of enforcing trademarks and the costs of protecting consumers from confusion falls on the mark owner. In considering the public interest, the mark holder's role in protecting consumers from confusion should be recognized. Moreover, the public has an interest in protecting trademarks because trademarks can facilitate business growth by ensuring that consumers can acquire goods and services they have previously purchased. Similarly, trademark owners are rewarded for selling goods and services of consistent quality when consumers are repeat customers. The trademark owner can also use its trademark as a "place-holder" in areas of likely expansion, which also facilitates business development.

The mark holder's role in enforcement is generally not a problem because the

"Parenthood" generally by the defendant, an opponent of abortion, to promote a book espousing an anti-abortion position). A court may apply the commercial use requirement to protect purely non-commercial uses, but the application of that test has led to some arguably inconsistent results. *Compare* *Bosley Med. Inst., Inc. v. Kremer*, 403 F.3d 672, 676-80 (9th Cir. 2005), *with* *People for the Ethical Treatment of Animals v. Doughney*, 263 F.3d 359, 365-66 (4th Cir. 2001).

131. *But see LANDES & POSNER, ECONOMIC STRUCTURE, supra* note 10, at 174 ("The fact that two goods have the same chemical formula does not make them of equal quality even to the most coolly rational consumer. The consumer will not be interested in the formula as such but in the actual manufactured product that he will be consuming, and he may therefore be willing to pay a premium for greater assurance that the product will actually be manufactured to the specifications of the formula.").

132. A similar concern is raised in situations involving marks that were used with patented products. When the patent expires, the mark holder is still able to benefit from the time it was able to exclusively sell the particular product or service because consumers strongly identify that mark with the formerly patented product. *See id.* at 314 ("[T]he patent monopoly accustoms doctors and patients to the name-brand product (trademark reinforcing patent), and when the patent expires and a generic substitute becomes available at much lower cost, they remain reluctant to substitute for the familiar brand an unknown (un)brand, which though certified as chemically identical may differ in some subtle way, perhaps involving quality control. The sales of the branded drug will fall, as the manufacturer in effect cedes the low-price segment of the market to the generics, but profit per unit will remain high, and aggregate profits, though smaller than before, will remain healthy. One study found only a gradual decline in market share upon patent expiration . . .").

133. *See generally* Viva R. Moffat, *Mutant Copyrights and Backdoor Patents: The Problem of Overlapping Intellectual Property Law Protection*, 19 BERKELEY TECH. L.J. 1473 (2004).

interests of mark holders and the public in preventing confusion typically overlap.¹³⁴ For example, a use that has the potential to confuse consumers is also one that a mark holder will want to prevent because it may result in a loss of sales. However, given the high value of marks, a mark holder is incentivized to overprotect its mark to ensure that no value from the mark is lost. This protection may well exceed that which is allowed by the consumer search cost theory.¹³⁵ Possible examples of courts validating this intuition may exist in initial interest confusion cases or cases finding trademark infringement by the purchase of keywords for use in search engines.¹³⁶ Those examples may be particularly troublesome to courts where there is a perceived clear intent to trade off the goodwill of the mark holder or deceive the public. Divorcing mark protection from the consumer search cost theory may allow for that protection, but only where there is some clear intent to deceive.¹³⁷

Another example of overprotection is disallowing uses of a mark that are inconsistent with the image the mark holder wants associated with that particular mark. Thus, any generally negative or inconsistent image or use of a mark may be chilled through the vigorous enforcement of trademark rights.¹³⁸ However, the use of a particular mark by a user or competitor may capture the exact message that the user wants to convey—one that could not be conveyed without that particular mark.¹³⁹ This problem highlights the concerns with free expression.¹⁴⁰

134. *See supra* note 80 and accompanying text.

135. *Ill. High Sch. Ass'n v. GTE Vantage Inc.*, 99 F.3d 244, 247 (7th Cir. 1996) (noting that the senior user cannot take a mark the public is using out of the public domain using and render the public speechless).

136. *Promatek Indus., Ltd. v. Equitrac Corp.*, 300 F.3d 808, 812-13 (7th Cir. 2002) (finding initial interest confusion and stating that “[w]hat is important is not the duration of the confusion, it is the misappropriation of Promatek’s goodwill. Equitrac cannot unring the bell.”); *Checkpoint Sys., Inc. v. Check Point Software Techs., Inc.*, 269 F.3d 270, 294-95 (3d Cir. 2001) (“Without initial interest protection, an infringer could use an established mark to create confusion as to a product’s source thereby receiving a ‘free ride on the goodwill’ of the established mark.”).

137. This is not the same as arguing that there should be an anti-free riding theory supporting trademark law. Rather, at the boundaries of the likelihood of confusion or dilution test, the trademark holder should be able to raise intent to deceive and protection of its mark in the consideration of the public interest in granting a preliminary or permanent injunction. This may be based on the public’s interest in fostering fair play in the marketplace. *See Doellinger, supra* note 89, at 823-24.

138. Paul Alan Levy, *The Trademark Dilution Revision Act—A Consumer Perspective*, 16 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1189, 1191-92 (2006) (“Trademark lawyers have become notorious for threatening litigation over the most questionable claims. Yet, threats of litigation, not to speak of receipt of an actual complaint, are highly intimidating because most individuals know nothing about trademark law and have no contact with experienced trademark counsel. And, if they consult a local lawyer, they are likely to learn how expensive and complicated the defense of trademark litigation can be.”).

139. Kozinski, *supra* note 2, at 973 (“Trademarks are often selected for their effervescent qualities, and then injected into the stream of communication with the pressure of a firehose by

Another public interest is the concern that trademark protection not be used to discourage new entrants to the market and harm competition by preventing competitors from using product design or packaging that may be needed to compete effectively.¹⁴¹ This concern has been expressed in the context of the functionality doctrine, although the application of that doctrine is unclear despite the Supreme Court's efforts to clarify it.¹⁴² Moreover, while the aesthetic functionality doctrine is supposed to protect against expansive uses of trademark law beyond those recognized by the consumer search cost theory,¹⁴³ some circuits have failed to adopt it¹⁴⁴ or apply it¹⁴⁵ in cases involving merchandising.

means of mass media campaigns. Where trademarks come to carry so much communicative freight, allowing the trademark holder to restrict their use implicates our collective interest in free and open communication.”). This problem is somewhat exemplified by the recent *Starbucks* case in which a parody defense to a dilution cause of action was not upheld because the allegedly infringing mark (a parody) was used as a mark itself. *Starbucks Corp. v. Wolfe’s Borough Coffee, Inc.*, 588 F.3d 97, 112 (2d Cir. 2009) (“As evident from the statutory language, Black Bear’s use of the Charbucks Marks cannot qualify under the parody exception because the Charbucks Marks are used ‘as a designation of source for [Black Bear’s] own goods[, i.e., the Charbucks line of coffee.]’”); *see also* *Louis Vuitton Malletier S.A. v. Haute Diggity Dog, LLC*, 507 F.3d 252, 266-67 (4th Cir. 2007) (recognizing that the TDRA does not extend the fair use defense to uses of the “parody as its own designation of source”).

140. *See Parks v. LaFace Records*, 329 F.3d 437, 449 (6th Cir. 2003) (“The public has at least as much interest in the free exchange of ideas as it does in avoiding misleading advertising. If Parks possesses a right to police the use of her name, even when that right can be exercised only to prevent consumer confusion, she has the means to restrict the public discourse to some extent.”).

141. Trademark law, particularly dilution, can deter new entrants to a mark to the extent it does not allow new entrants to use the marks of established competitors. *See* Laura R. Bradford, *Emotion, Dilution, and the Trademark Consumer*, 23 BERKELEY TECH. L.J. 1227, 1297 (2008) (“The ability of consumers to rely on the credibility of established brands offers real benefits. On the other hand, the ability of new producers to reference stronger marks eases barriers to entry.”); *see generally* Ghosh, *supra* note 22.

142. *See TrafFix Devices, Inc. v. Mktg. Displays, Inc.*, 532 U.S. 23, 32-33 (2001); *Eppendorf-Netheler-Hinz GMBH v. Ritter GMBH*, 289 F.3d 351, 355-56 (5th Cir. 2002); *Valu Eng’g, Inc. v. Rexnord Corp.*, 278 F.3d 1268, 1277-78 (Fed. Cir. 2002).

143. LANDES & POSNER, ECONOMIC STRUCTURE, *supra* note 10, at 200 (“[C]ourts deny trademark protection for the attractive feature only if it is indispensable to the marketing of the product, that is, roughly speaking, only if the trademark owner would obtain a product monopoly if he could exclude others from copying the feature. This approach permits legal protection for attractive colors that serve as identifiers, such as the color pink for insulation but not the color yellow for tennis balls or the color brown for peanut butter—or the color yellow for margarine.”).

144. *See, e.g.*, Bd. of Supervisors for La. State Univ. Agric. & Mech. Coll. v. Smack Apparel Co., 550 F.3d 465, 487-88 (5th Cir. 2008).

145. *See, e.g.*, Au-Tomotive Gold, Inc. v. Volkswagen of Am., Inc., 457 F.3d 1062 (9th Cir. 2006). However, a recent Ninth Circuit case applied this doctrine where the alleged trademark owner was using the mark as a product, not a trademark. *See Fleischer Studios, Inc. v. A.V.E.L.A., Inc.*, No. 09-56317, 2011 WL 631499, at *7 (9th Cir. Feb. 23, 2011).

There are several other potential public interests that implicate users' interests. The protection and use of trademarks leads to the development of new symbols and meanings that enrich our lives and facilitate self-definition and expression.¹⁴⁶ This public interest seems to comport with consumer search cost theory, but in cases of dilution without confusion, there may be some harm to consumers not properly based on a concern with consumer search costs.¹⁴⁷ Moreover, the public interest in "the right to cultural participation, including . . . [the] cultural rights of specific groups" could be taken into account¹⁴⁸ and viewed as an interest in the actual mark users.¹⁴⁹

Accordingly, the consumer search cost theory leads to broader trademark rights because the theory is primarily concerned with viewing trademarks from the consumer's perspective, and producers are in (at least partial) control of how consumers view their marks.¹⁵⁰ The theory also does not adequately address other important values implicated by trademark protection. Thus, the other interests—from the user, consumer, and competitor perspective—must be taken into account and balanced, and the question is when and how we allow these interests to be considered.¹⁵¹

146. See Dreyfuss, *supra* note 2, at 397 ("[I]deograms that once functioned solely as signals denoting the source, origin, and quality of goods, have become products in their own right, valued as indicators of the status, preferences, and aspirations of those who use them."); Swann et al., *supra* note 45, at 796.

147. Professor Landes and Judge Posner contend that one economic argument for antidilution laws is that people "advertise themselves (much as sellers advertise their goods) by wearing clothes, jewelry, or accessories that tell the world that they are people of refined or flamboyant taste or high income." LANDES & POSNER, ECONOMIC STRUCTURE, *supra* note 10, at 208. They state that "[i]f others can buy and wear cheap copies, the signal given out by the purchasers of the originals is blurred." *Id.* at 208-09.

148. See Megan M. Carpenter, *Trademarks and Human Rights: Oil and Water? Or Chocolate and Peanut Butter?*, 99 TRADEMARK REP. 892, 929 (2009) ("Within the framework of human rights, indigenous peoples' rights must be balanced alongside the rights of trademark owners, and, as human rights discourse begins to consider issues of trademark rights, those rights may be antagonistic to the ability of indigenous groups to enjoy certain cultural rights respecting cultural participation, identity, and self-determination.").

149. Notably, the consumer search cost theory does not explain the prohibition on scandalous and disparaging marks.

150. See Kratzke, *supra* note 82, at 209; McKenna, *Normative Foundations*, *supra* note 15, at 1916; McKenna, *Trademark Use*, *supra* note 14, at 773. Professors Dogan and Lemley would have the trademark "use" doctrine serve to limit trademark infringement cases where the defendant does not use the mark as a trademark, thus insulating alleged infringers from liability for using the marks for "speech-oriented" objectives. See Dogan & Lemley, *Search Costs*, *supra* note 65, at 808-11.

151. Professors Dinwoodie and Janis argue that these or similar interests can be taken into account through the creation of defenses. See Graeme B. Dinwoodie & Mark D. Janis, *Confusion over Use: Contextualism in Trademark Law*, 92 IOWA L. REV. 1597, 1662 (2007) [hereinafter Dinwoodie & Janis, *Contextualism*] (stating that courts should create and modify defenses to take

B. The Structure of Trademark Law

The structure of trademark law provides an incentive for mark holders to enforce their rights vigorously.¹⁵² This is largely based on the fact that much of the legal protection for trademark law is based upon consumer perception.¹⁵³ Consumer perception impacts whether marks are entitled to protection in the first instance or can ever be protected; the scope of protection of a mark in infringement proceedings; the availability and scope of an antidilution action; whether a defense may apply; and whether a mark has been abandoned.¹⁵⁴

1. *Distinctiveness and Secondary Meaning.*—Whether a particular word, design, or composite mark is entitled to federal protection is dependent upon where the mark falls on the spectrum of distinctiveness.¹⁵⁵ The overarching question is whether the mark identifies and distinguishes the goods and services of one mark holder from those of another. This analysis rests upon how a consumer may perceive the mark based on the good or service the mark is used with, and it turns on distinctiveness. A mark that is inherently distinctive is automatically entitled to legal protection.¹⁵⁶ Inherently distinctive marks may include fanciful marks, completely made-up words; arbitrary marks, words that bear no relationship to the goods and services with which they are used; and suggestive marks, marks that suggest, but do not describe, a characteristic or quality of the good or service.¹⁵⁷ Marks that have a primary meaning in

into account values such as comparative advertising, free speech, and anticompetition concerns); *see also* Dinwoodie, *supra* note 24, at 99 (arguing for the development of defenses that take into account values other than those driven purely by consumer confusion concerns). Professor Grynberg argues that formalist courts are unlikely to expand or develop defenses to address the concerns raised by Professors Dinwoodie and Janis. *See* Grynberg, *supra* note 24, at 901-02. Professor Grynberg posits that the infringement standard with a materiality requirement may limit some of trademark law's expansions. *Id.* at 967-69. Although some of these interests may be taken into account in an infringement analysis, they may not be considered at an early enough procedural stage to deter the chilling effect of the exercise of these interests. I argue that these interests should be taken into account at the preliminary injunction stage.

152. *See* Ill. High Sch. Ass'n v. GTE Vantage Inc., 99 F.3d 244, 246 (7th Cir. 1996) ("What matters is that a trademark is not nearly so secure an entitlement as a property right. It is mainly just a designation of source, and dies when it ceases to designate, for whatever reason other than the culpable conduct of the defendant." (internal citations omitted)).

153. *See* McKenna, *Normative Foundations*, *supra* note 15, at 1916; McKenna, *Trademark Use*, *supra* note 14, at 812 ("Indeed, virtually every distraction courts and the Trademark Office draw between protectable and unprotectable matter depends on consumer understanding.").

154. *See* McKenna, *Trademark Use*, *supra* note 14, at 812.

155. Abercrombie & Fitch Co. v. Hunting World, Inc., 537 F.2d 4, 9-10 (2d Cir. 1976); *see also* Zatarains, Inc. v. Oak Grove Smokehouse, Inc., 698 F.2d 786, 790-92 (5th Cir. 1983), abrogated by KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc., 543 U.S. 111 (2004).

156. *Abercrombie & Fitch Co.*, 537 F.2d at 9-10; *see* Zatarains, 698 F.2d at 790-92.

157. *Abercrombie & Fitch Co.*, 537 F.2d at 9-10; *see* Zatarains, 698 F.2d at 790-92.

connection with the good or service and describe the good or service are protected on a showing of acquired distinctiveness—so-called “secondary meaning.”¹⁵⁸ Marks that identify the particular good or service (or class of goods or services) are considered generic and are not protectable.¹⁵⁹

For a mark to acquire distinctiveness, a mark holder needs to use the mark in connection with the good or service so that consumers will perceive the mark as indicating source and distinguishing one person’s goods from another (as opposed to merely describing the good or service).¹⁶⁰ Thus, the mark holder has an interest in ensuring that the mark develops secondary meaning, and the holder can manipulate consumer perception through advertising. The mark holder can also attempt to acquire secondary meaning and control consumer perception by ensuring that third parties do not use a similar mark in connection with similar goods or services. Therefore, even though a mark holder may not have legally protected rights, it could assert “rights” through cease and desist letters and thereby discourage the use of similar marks, eventually helping the mark holder establish trademark rights in the first instance.¹⁶¹ Accordingly, the determination of whether a mark should initially receive protection by a finding of acquired distinctiveness provides an incentive for mark holders to police and prevent third party use of their marks.¹⁶²

158. *Abercrombie & Fitch Co.*, 537 F.2d at 9-10; *see Zatarains*, 698 F.2d at 790-92.

159. *Abercrombie & Fitch Co.*, 537 F.2d at 9-10; *see Zatarains*, 698 F.2d at 790-92.

160. *Abercrombie & Fitch Co.*, 537 F.2d at 9-10; *see Zatarains*, 698 F.2d at 790-92.

161. In a recent opinion, the Eastern District of California considered and relied upon evidence of the plaintiff’s policing efforts (cease and desist letters) to find no knowing false misrepresentation of fact to the Patent and Trademark Office based on the plaintiff’s affidavit, which claimed “substantially exclusive and continuous” use of a mark for five years to obtain a presumption of acquired distinctiveness, even with evidence of two other instances of third party use. *Salu, Inc. v. Original Skin Store*, No. Civ-S-08-1035, 2010 WL 1444617, at *4 (E.D. Cal. Apr. 12, 2010). The court stated:

[Plaintiff] Salu presents evidence that . . . [it] has continually contacted anyone who it feels is infringing in order to protect its mark. It has sent out over 300 cease and desist letters to alleged infringers in the last couple of years alone. Salu claims that with the exception of ESKINSTORE, this litigation, and one other case that settled out of court, “every other infringer receives [the] letter and stops infringing on [the] trademark.”

Id. at *2 (internal citations omitted).

162. See *Yellow Cab Co. of Sacramento v. Yellow Cab of Elk Grove, Inc.*, 419 F.3d 925, 930 (9th Cir. 2005) (deciding whether exclusive usage of a trademark was probative of secondary meaning); *DeGidio v. W. Grp. Corp.*, 355 F.3d 506, 513-14 (6th Cir. 2004) (finding that wide use of a term weighed against a finding of secondary meaning); *Int’l Bancorp, LLC v. Societe des Bains de Mer et du Cercle des Etrangers a Monaco*, 329 F.3d 359, 370-71 (4th Cir. 2003) (approving district court’s reliance in part on “continuous, if not exclusive use” of mark to support a finding of secondary meaning); *Commerce Nat’l Ins. Servs., Inc. v. Commerce Ins. Agency, Inc.*, 214 F.3d 432, 437 (3d Cir. 2000) (stating that one of the factors to determine secondary meaning is exclusivity of use of a mark); *Petro Stopping Ctrs., L.P. v. James River Petroleum, Inc.*, 130 F.3d 88, 93-94 (4th Cir. 1997) (approving district court’s reliance on third party registrations and usages

In determining whether some types of trade dress or logos receive legal protection, some courts find the spectrum of distinctiveness unhelpful and rely upon the *Seabrook* test.¹⁶³ The *Seabrook* test provides:

In determining whether a design is arbitrary or distinctive this court has looked to [1] whether it was a ‘common’ basic shape or design, [2] whether it was unique or unusual in a particular field, [3] whether it was a mere refinement of a commonly-adopted and well-known form of ornamentation for a particular class of goods viewed by the public as a dress or ornamentation for the goods, or [4] whether it was capable of creating a commercial impression distinct from the accompanying words.¹⁶⁴

Each factor of this test focuses on the perception of the consumer, and all factors may require examination of the use of the mark by third parties.¹⁶⁵ Thus, the test pushes mark holders to litigate to ensure that their marks are initially protected and continue to receive protection.¹⁶⁶

2. *Genericness and Genericide.*—The final category on the spectrum of distinctiveness is generic marks.¹⁶⁷ A mark may be considered generic at the outset of use by the mark holder, or it may eventually become the generic term for a once-protectable mark in connection with a specific good or service.¹⁶⁸ The former is generally referred to as “genericness,” and the latter is “genericide.”

The question of whether a mark is generic is determined from the perspective of the consumer.¹⁶⁹ Generic marks cannot receive protection because of the strong interest in allowing competitors to use a particular name to call their

of the term “Petro” to support a finding that the mark was descriptive and weak as part of strength of mark analysis); Spraying Sys. Co. v. Delavan, Inc., 975 F.2d 387, 393 (7th Cir. 1992) (finding that evidence of third party use was probative of secondary meaning); Alphaville Design, Inc. v. Knoll, Inc., 627 F. Supp. 2d 1121, 1128-1130 (N.D. Cal. 2009) (considering, in determining secondary meaning, evidence of failure to police third parties’ use of marks as relevant to whether a purported trademark owner exclusively used marks); Simon & Schuster, Inc. v. Dove Audio, Inc., 936 F. Supp. 156, 163 (S.D.N.Y. 1996) (finding the existence of secondary meaning supported by sufficient allegations, including successful policing of mark).

163. *Seabrook Foods, Inc. v. Bar-Well Foods Ltd.*, 568 F.2d 1342 (C.C.P.A. 1977).

164. *Id.* at 1344 (internal citations omitted).

165. See *Amazing Spaces, Inc. v. Metro Mini Storage*, 608 F.3d 225, 247 (5th Cir. 2010) (finding logo not inherently distinctive under *Seabrook* analysis based in part on widespread use of similar logo).

166. Depending on when the court determines distinctiveness under *Seabrook*, the enforcement of a mark may be relevant to the strength of the mark, not initial protection.

167. *Zatarains, Inc. v. Oak Grove Smokehouse, Inc.*, 698 F.2d 786, 790 (5th Cir. 1983), abrogated by *KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc.*, 543 U.S. 111 (2004); *Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4, 9 (2d Cir. 1976).

168. *Schwan’s IP, LLC v. Kraft Pizza Co.*, 460 F.3d 971, 976 (8th Cir. 2006).

169. See *Yellow Cab Co. of Sacramento v. Yellow Cab of Elk Grove, Inc.*, 419 F.3d 925, 929 (9th Cir. 2005).

products by that particular product name—which, if denied to competitors, harms consumers.¹⁷⁰ For example, a purported trademark holder would not be able to acquire protection for the word “Apple” used in connection with the sale of apples.¹⁷¹ Thus, parties that hope to acquire trademark protection have an incentive to choose words that are not the particular names of existing products or services (or product or service classes). In determining whether a particular mark is initially generic in connection with a certain good or service, courts may consider the extent of third party use of the mark.¹⁷² A special problem arises when mark holders choose a mark for a product type that did not previously exist—for example, where the holder has created a new product class.¹⁷³ Each of these mark holders runs the risk that its mark is the generic name for that product. Because whether marks are protected is determined from the consumer’s perspective, mark owners have an interest in ensuring that their marks are not initially deemed generic. Thus, stopping third party mark usage

170. *Am. Cyanamid Corp. v. Connaught Labs., Inc.*, 800 F.2d 306, 308 (2d Cir. 1986) (“Were the first user of a generic or descriptive term . . . able to exclude later entrants from use of that term, the former would be able not only to identify itself as the maker of the . . . [product] and to capitalize on whatever good will it has built up—legitimate purposes of trademark protection—but also to impair the ability of competitors to describe their products . . . [in the same way]—a wholly counterproductive result so far as consumers are concerned.”); *Abercrombie & Fitch Co.*, 537 F.2d at 9.

171. *Dogan & Lemley, Search Costs, supra* note 65, at 809.

172. See *Schwan's*, 460 F.3d at 974-75 (considering evidence of use of mark by several competitors in finding the mark generic); *Nartron Corp. v. STMicroelecs., Inc.*, 305 F.3d 397, 406 (6th Cir. 2002) (“[Defendant] provided testimony that manufacturers, customers, suppliers, vendors, and other in the semiconductor industry, including trade and technical press, use the term ‘smart power’ generically to mean power devices that have control circuits.”); *Nat'l Envelope Corp. v. Am. Pad & Paper Co. of Del.*, No. 06 Civ. 12988, 2009 WL 5173920, at *3 (S.D.N.Y. Dec. 30, 2009) (noting that courts “may consider all relevant evidence, including . . . uncontested usage of the mark by competitors to describe their products”); *FragranceNet.com v. Les Parfums, Inc.*, 672 F. Supp. 2d 328, 333 (E.D.N.Y. 2009) (finding that evidence of genericness included uncontested uses by competitors to describe their goods and services); *CG Roxane LLC v. Fiji Water Co.*, 569 F. Supp. 2d 1019, 1028 (N.D. Cal. 2008) (“[D]efendant has still proven that at least 70 competitors use the phrasé ‘bottled at the source.’ This evidence of use by competitors strongly suggests that the mark is generic”); *Premier Nutrition, Inc. v. Organic Food Bar, Inc.*, 475 F. Supp. 2d 995, 1002 (C.D. Cal. 2007) (finding that a party challenging a mark as generic raised an inference of genericness with proof that competitors used the mark to refer to a category of products); *Classic Foods Int'l Corp. v. Kettle Foods, Inc.*, 468 F. Supp. 2d 1181, 1190 (C.D. Cal. 2007) (“Federal courts also view usage of the term by competitors in the industry as strong evidence of how the public perceives the term [in genericness inquiries].”).

173. See generally *Genesee Brewing Co. v. Stroh Brewing Co.*, 124 F.3d 137 (2d Cir. 1997); *A.J. Canfield Co. v. Honickman*, 808 F.2d 291 (3d Cir. 1986); 2 J. THOMAS McCARTHY, *McCARTHY ON TRADEMARKS AND UNFAIR COMPETITION* § 12:26 (4th ed. 2008) (describing how parties who create a new product should choose two words to market the product: one word to use as the product’s generic name, and the other word as the product’s trademark).

early is beneficial to mark owners.¹⁷⁴

Protectable marks can become generic over time—particularly fanciful marks. In determining whether there is so-called “genericide,” courts will examine the use of the mark by consumers and even consider the policing efforts of mark holders against third parties.¹⁷⁵ If the mark holder does not adequately control the use of its mark, the mark holder may lose trademark protection.¹⁷⁶

174. Some courts may determine whether a mark is generic by only considering evidence of genericness at the time the purported mark holder entered the market. *Compare Classic Foods*, 468 F. Supp. 2d at 1188 n.10, 1190-91 (considering competitor usage of a term before the alleged mark holder entered the market), *with Schwan's*, 460 F.3d at 974-75 (considering evidence of competitor use of a mark before and after the alleged mark holder entered the market in apparent genericness analysis).

175. *Freecycle Network, Inc. v. Oey*, 505 F.3d 898, 905-06 (9th Cir. 2007) (noting that policing a mark may prevent genericide); *Ty Inc. v. Softbelly's, Inc.*, 353 F.3d 528, 531 (7th Cir. 2003) (finding that evidence of the BEANIES mark's primary significance included Ty's policing “the use of ‘Beanie(s)’ vigorously by filing lawsuits, sending cease and desist letters, and opposing trademark applications for the word or its cognates”); *BellSouth Corp. v. DataNational Corp.*, 60 F.3d 1565, 1570 (Fed. Cir. 1995) (finding that a mark may become generic because of a failure to police third party uses); *Malaco Leaf, AB v. Promotion in Motion, Inc.*, 287 F. Supp. 2d 355, 364 (S.D.N.Y. 2003) (finding a mark generic based on extensive third party use and failure to police); *Ty, Inc. v. Jones Grp., Inc.*, 98 F. Supp. 2d 988, 994 (N.D. Ill. 2000) (“[G]enericness can be shown by generic use of the mark by competitors which has not been contested by the plaintiff. Plaintiff contends that the mark is not generic based on the fact that it has rigorously policed the use of the mark by competitors.”), *aff'd*, 237 F.3d 891 (7th Cir. 2001); *Brandwynne v. Combe Int'l, Ltd.*, 74 F. Supp. 2d 364, 381 (S.D.N.Y. 1999) (finding that failure to police third party usage of a purported mark is relevant to whether the mark is generic); *E.I. DuPont de Nemours & Co. v. Yoshida Int'l, Inc.*, 393 F. Supp. 502, 527 (E.D.N.Y. 1975) (for trademark protection program, considering policing misuse of a mark in finding the mark non-generic); *see also Deven R. Desai & Sandra L. Rierson, Confronting the Genericism Conundrum*, 28 CARDOL. REV. 1789, 1833 (2007) (“[T]he current doctrine's insistence that trademark holders ‘police’ against noncompetitive, noninfringing uses of their marks leads to overly aggressive trademark enforcement activities, which are inefficient and, perhaps more importantly, may have the undesirable effect of stifling the public’s ability to use language as it sees fit.”); *Ralph H. Folsom & Larry L. Teply, Trademarked Generic Words*, 89 YALE L.J. 1323, 1354 (1980) (“[I]n some cases the fact that a trademark holder has engaged in policing efforts has been viewed as evidence in favor of continuing trademark rights. . . . We argue . . . that evidence of extensive policing efforts should be weighted *negatively* in determinations of genericness.” (emphasis in original) (internal citations omitted)). However, trademark holders may have little success in preventing generic usage by the media and dictionaries. *See 2 MCCARTHY, supra* note 173, at § 12:28 (citing *Freecycle*, 505 F.3d at 898; *Ty Inc. v. Perryman*, 306 F.3d 509 (7th Cir. 2002); *Ill. High Sch. Ass'n v. GTE Vantage Inc.*, 99 F.3d 244 (7th Cir. 1996)).

176. *E.g., King-Seeley Thermos Co. v. Aladdin Indus., Inc.*, 321 F.2d 577, 579 (2d Cir. 1963). For example,

[a]lthough “Thermos” was generally recognized in the trade as a trademark, the corporation did police the trade and notified those using “thermos” in a descriptive

3. *Strength of the Mark.*¹⁷⁷—In a field related to the spectrum of distinctiveness, a very relevant concern to a mark holder is the strength of the mark—a particular mark’s degree of distinctiveness.¹⁷⁸ The strength of a mark can determine the scope of protection it receives.¹⁷⁹ The stronger a mark is, the more likely that infringement may be found, even against a mark that is used on goods that are only slightly related. The strength of a mark is demonstrated using either direct or circumstantial evidence. Direct evidence may involve customer surveys, while circumstantial evidence may include the amount of advertising expenditures or sales. The strength of the mark can also be assessed by examining where the mark falls on the spectrum of distinctiveness.¹⁸⁰ For

sense that it was a trademark. It failed, however, to take affirmative action to seek out generic uses by non-trade publications and protested only those which happened to come to its attention. Between 1923 and the early . . . [1950s] the generic use of “thermos” had grown to a marked extent in non-trade publications and by the end of this period there was wide-spread use by the unorganized public of “thermos” as a synonym for “vacuum insulated.” The court concluded that King-Seeley had failed to use due diligence to rescue “Thermos” from becoming a descriptive or generic term.

Id.; see also *Magic Wand, Inc. v. RDB, Inc.*, 940 F.2d 638, 641 (Fed. Cir. 1991) (considering evidence of third party usage in finding that a mark was not generic to the relevant class of purchasers); *Murphy Door Bed Co. v. Interior Sleep Sys., Inc.*, 874 F.2d 95, 101 (2d Cir. 1989) (noting that the district court’s finding of non-genericness relied on policing efforts of mark claimant, the court found that policing efforts were not helpful where the mark had already entered the public domain); *Pilates, Inc. v. Current Concepts, Inc.*, 120 F. Supp. 2d 286, 300 (S.D.N.Y. 2000) (finding that policing activity was not relevant once a mark had become generic). Notably, it is ultimately the consumer’s perception that matters and not expenditures on advertising or policing, although these are relevant categories of evidence to determine genericide. See *Magic Wand*, 940 F.2d at 641 (“[W]hether a term is entitled to trademark status turns on how the mark is understood by the purchasing public.”). For further discussion of genericism, see Jerre B. Swann, *Genericism Rationalized*, 89 TRADEMARK REP. 639 (1999).

177. Notably, the likelihood of confusion factor is important because it, along with the strength of the mark factor, can expand trademark protection beyond competing goods and services. Some courts state that if the related good or service market is one in the natural zone of expansion of the trademark holder, that factor favors the trademark holder. See, e.g., *Westchester Media v. PRL USA Holdings, Inc.*, 214 F.3d 658, 666-67 (5th Cir. 2000). The natural zone of expansion is judged from the perspective of the consumer. *Id.* at 666.

178. 2 MCCARTHY, *supra* note 173, at § 11:75. Interestingly, Professor Beebe, in an empirical study, found that there was a “good correlation between inherent strength and success in the multifactor test” in infringement actions—although perhaps not as strong as believed. Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 CAL. L. REV. 1581, 1634 (2006).

179. See *One Indus., LLC v. Jim O’Neal Distrib., Inc.*, 578 F.3d 1154, 1164 (9th Cir. 2009) (observing that “[t]rademark law offers greater protection to marks that are ‘strong’”) (quoting E. & J. Gallo Winery v. Gallo Cattle Co., 967 F.2d 1280, 1291 (9th Cir. 1992)).

180. See *Utah Lighthouse Ministry v. Found. for Apologetic Info. & Research*, 527 F.3d 1045, 1056 (10th Cir. 2008) (“Strength consists of both conceptual strength, which refers to the placement

example, a mark that is considered arbitrary is generally stronger than a merely descriptive mark, which is unprotectable without a showing of acquired distinctiveness. If a descriptive mark has acquired distinctiveness, the degree of strength of that mark can be a very important question. A descriptive mark with acquired distinctiveness or a suggestive mark may generally possess a low level of strength and thus receive a limited scope of protection. However, evidence of large amounts of sales or advertising expenditures can increase the strength of the mark.¹⁸¹ Again, the mark holder controls the amount and type of advertising it uses.

Another measure of strength of the mark is the amount of third party usage of the particular mark in connection with the same or similar goods.¹⁸² A large amount of third party usage may indicate that the mark is weak and deserves a minimal scope of protection.¹⁸³ Mark holders are thus incentivized to aggressively police their marks. Professor J. Thomas McCarthy has noted this effect:

The law imposes on trademark owners the duty to be pro-active and to police the relevant market for infringers. If the trademark owner is quiescent and tolerates the encroachment of infringers, it will find that its trademark asset has “eroded” and “shrunken” because the strength of its mark as a distinctive and distinguishing symbol has been diminished by the presence of similar marks. . . . [T]he trademark owner must anticipate constant “maintenance” of its trademark property by policing and enforcing the exclusivity of its trademark symbol in the marketplace.¹⁸⁴

Indeed, numerous courts have stated that failure to police one’s mark or extensive third party usage of a mark will result in a “weaker” mark or a smaller scope of protection, or that policing a mark may lead to a “stronger” mark.¹⁸⁵ The courts’

of the mark along the distinctiveness spectrum, and commercial strength, which refers to the marketplace recognition value of the mark.”).

181. *Brookfield Commc’ns, Inc. v. W. Coast Entm’t Corp.*, 174 F.3d 1036, 1058 (9th Cir. 1999); *Levis Strauss & Co. v. Blue Bell, Inc.*, 632 F.2d 817, 821-22 (9th Cir. 1980).

182. *Freedom Sav. & Loan Ass’n v. Way*, 757 F.2d 1176, 1182 (11th Cir. 1985) (noting that “the strength of the mark depends on the extent of third party usage”).

183. *One Indus.*, 578 F.3d at 1164-65 (stating that third party uses weaken a distinctive mark); *Citizens Fin. Grp., Inc. v. Citizens Nat’l Bank of Evans City*, 383 F.3d 110, 123 (3d Cir. 2004) (“[A]s a general rule, widespread use of even a distinctive mark may weaken the mark.”). *But see Skechers U.S.A., Inc. v. Vans, Inc.*, No. CV 07-01703, 2007 WL 4181677, at *5 (C.D. Cal. Nov. 20, 2007) (holding that third party usage of a checkerboard mark on shoes would cause consumers to carefully distinguish between the mark holder’s mark and third party uses and would thus “significantly diminish[] the likelihood of confusion”).

184. 2 MCCARTHY, *supra* note 173, at § 11:91.

185. *Morningside Grp. Ltd. v. Morningside Capital Grp. L.L.C.*, 182 F.3d 133, 139 (2d Cir. 1999) (“[T]he successful policing of a mark adds to its strength to the extent that it prevents weakening of the mark’s distinctiveness in the relevant market.” (internal citation omitted));

Streetwise Maps, Inc. v. VanDam, Inc., 159 F.3d 739, 743-44 (2d Cir. 1998); Petro Stopping Ctrs., L.P. v. James River Petroleum, Inc., 130 F.3d 88, 93-94 (4th Cir. 1997) (approving district court's reliance on third party registrations and usages of the term "Petro" to support a finding that the mark was descriptive and weak as part of strength of mark analysis); First Sav. Bank v. First Bank Sys., Inc., 101 F.3d 645, 653-54 (10th Cir. 1996) ("The greater the number of identical or more or less similar marks already in use on different kinds of goods, the less the likelihood of confusion between any two specific uses of the weak mark."); Lang v. Ret. Living Publ'g Co., 949 F.2d 576, 581 (2d Cir. 1991) (relying on third party usage of parts of a mark in finding the mark weak); Sun Banks of Fla., Inc. v. Sun Fed. Sav. & Loan Ass'n, 651 F.2d 311, 316 (5th Cir. 1981) ("[W]e find the extensive third-party use of the word 'Sun' impressive evidence that there would be *no* likelihood of confusion between Sun Banks and Sun Federal." (emphasis in original)); Amstar Corp. v. Domino's Pizza, Inc., 615 F.2d 252, 260 (5th Cir. 1980) ("The third-party uses and registrations . . . merely limit the protection to be accorded plaintiff's mark outside the uses to which plaintiff has already put its mark."); Alaven Consumer Healthcare, Inc. v. DrFloras, LLC, No. 1:09-CV-705-TWT, 2010 WL 481205, at *2 (N.D. Ga. Feb. 4, 2010) ("Although suggestive, the trademark is weakened by the frequency with which the prefix 'Dr.' is used in other trademark registrations for related products."), *aff'd in part by* 399 F. App'x 545 (11th Cir. 2010); Am. ORT, Inc. v. ORT Israel, No. 07 CV 2332, 2007 WL 2049733, at *7 (S.D.N.Y. July 17, 2007) (holding that policing of mark supported a finding of a strong mark); Century 21 Real Estate LLC v. Century Ins. Grp., No. CIV 03-0053-PHX-SMM, 2007 WL 484555, at *11 (D. Ariz. Feb. 9, 2007) ("Given the Court's findings regarding the inherent strength of . . . [Century 21's] Marks in the real estate industry, the weakness of the Marks in the insurance industry, and weakness resulting from third-party use of the term 'Century' in real estate and insurance industries, the strength factor favors . . . [Century 21] but only slightly." (citation omitted)), *aff'd*, 300 F. App'x 527 (9th Cir. 2008); U.S. Conference of Catholic Bishops v. Media Research Ctr., 432 F. Supp. 2d 616, 626 (E.D. Va. 2006) ("Plaintiff's mark . . . has been further diluted by significant third-party usage, including third-party usage in Plaintiff's exact line of business."); Children's Med. Ctr. of Dall. v. Columbia Hosp. at Med. City Dall., No. 3-04-CV-2436-BD, 2006 WL 616000, at *5 (N.D. Tex. Mar. 10, 2006) (admitting expert testimony that a mark was weak based on extensive third party use of the mark); Renaissance Greeting Cards, Inc. v. Dollar Tree Stores, Inc., 405 F. Supp. 2d 680, 692 (E.D. Va. 2005) ("The common use of a given trademark is still considered an important factor in considering the strength of a mark."), *aff'd*, 227 F. App'x 239 (4th Cir. 2007); Gateway, Inc. v. Companion Prods., Inc., No. Civ. 01-4096-KES, 2003 WL 22508907, at *6 (D.S.D. Aug. 19, 2003) (relying on evidence of cease and desist letters and litigation to find a mark strong); M & G Elecs. Sales Corp. v. Sony Kabushiki Kaisha, 250 F. Supp. 2d 91, 102 (E.D.N.Y. 2003) (noting extensive third-party use and lack of evidence of policing in finding mark weak); Sterling Acceptance Corp. v. Tommark, Inc., 227 F. Supp. 2d 454, 462 (D. Md. 2002) ("Significant evidence in the record here of third-party use of the word 'Sterling' in the financial services industry and in other trademark registrations leads to the conclusion that plaintiff's mark is weak and does not acquire significant secondary meaning."), *aff'd*, 91 F. App'x 880 (4th Cir. 2004); Checkpoint Sys., Inc. v. Check Point Software Techs., Inc., 104 F. Supp. 2d 427, 459 (D.N.J. 2000) ("Checkpoint's active program of policing its trademark is further proof of the strength of its mark."), *aff'd*, 269 F.3d 270 (3d Cir. 2001); Quantum Fitness Corp. v. Quantum LifeStyle Ctrs., LLC, 83 F. Supp. 2d 810, 820 (S.D. Tex. 1999) ("Another factor in determining the strength of a mark in the marketplace is the extent of third-party usage of similar marks."); Gideons Int'l, Inc. v. Gideon 300 Ministries, Inc., 94 F.

consideration of evidence of a mark's strength by examining its usage outside the product and services market of the mark holder is particularly troubling because it provides an incentive for mark holders to police potentially infringing uses outside their markets.¹⁸⁶ This policing activity may allow mark holders to expand their rights beyond their usual product or service market to other related markets.

4. *Dilution*.—Under the TDRA, only famous marks are protected.¹⁸⁷ In order for a mark to be deemed famous, it must be widely known to the general consuming public.¹⁸⁸ This standard is used to ensure that only truly powerful

Supp. 2d 566, 583 (E.D. Pa. 1999) (“The fact that The Gideons have taken steps over the years to enforce its marks, including sending cease and desist letters to potential infringers . . . as a part of a regular and systematic practice . . . is also evidence of the strength of The Gideons’ marks as accorded by third parties.”); Schieffelin & Co. v. Jack Co. of Boca, 850 F. Supp. 232, 243 (S.D.N.Y. 1994) (“Schieffelin’s consistent effort to prevent use of its marks in a manner that would either confuse consumers or undermine the goodwill associated with its brands is, therefore, further proof of the strength of plaintiff’s marks.”); Hester Indus., Inc. v. Wing King, Inc., No. 1:91-CV-2644-RHH, 1992 WL 200129, at *3 (N.D. Ga. Mar. 26, 1992) (relying in part on the plaintiff’s vigorous policing efforts in deeming the mark strong), *aff’d*, 979 F.2d 1539 (11th Cir. 1992); Berkshire Fashions, Inc. v. Sara Lee Corp., 725 F. Supp. 790, 795 (S.D.N.Y. 1989) (“The evidence offered by Berkshire of third party use of energy evoking marks, assertedly to establish dilution of the strength of Sara Lee’s mark, and of putative inadequate policing of its mark by Sara Lee, is rejected as marginal and not sufficiently comprehensive with respect to market use, presence and inter-play of the respective brand products.”), *aff’d*, 904 F.2d 33 (2d Cir. 1990); E.I. DuPont de Nemours & Co. v. Yoshida Int’l, Inc., 393 F. Supp. 502, 512 (E.D.N.Y. 1975) (“[S]trength is primarily a question of degree, an amorphous concept with little shape or substance when divorced from the mark’s commercial context, including an appraisal of the owner’s policing efforts to ensure that whatever distinctiveness or exclusivity has been achieved is not lost through neglect, inattention, or consent to infringing use.”). In the Federal Circuit, an analysis of third party usage is a separate factor in examining whether marks are likely to be confused under Section 2(d) of the Lanham Act. *See In re Mighty Leaf Tea*, 601 F.3d 1342, 1346-47 (Fed. Cir. 2010).

186. Bd. of Supervisors for La. State Univ. Agric. & Mech. Coll. v. Smack Apparel Co., 550 F.3d 465, 479 (5th Cir. 2008) (“All third-party use of a mark, not just use in the same industry as a plaintiff, may be relevant to whether a plaintiff’s mark is strong or weak.”); A & H Sportswear, Inc. v. Victoria’s Secret Stores, Inc., 237 F.3d 198, 223 (3d Cir. 2000) (“Although the wide use of a term within the market at issue is more probative than wide use of a term in other markets, the extensive use of the term in other markets may also have a weakening effect on the strength of the mark.” (internal citation omitted)); *see Sterling Acceptance Corp.*, 227 F. Supp. 2d at 462 (“[E]vidence disclosed by the Patent and Trademark Office’s online database show eighteen registrations in the financial services class which include the word ‘Sterling’ and over three hundred and fifty registrations in other classes which include the word ‘Sterling.’”). *But see Lahoti v. Vericheck, Inc.*, No. C06-1132JLR, 2010 WL 1473976 (W.D. Wash. Apr. 9, 2010) (deeming a mark not weak based on third party usage, mostly in unrelated service markets).

187. 15 U.S.C. § 1125(c)(1) (2006). The dilution concept first appeared in an article by Frank Schechter. *See* Frank I. Schechter, *The Rational Basis of Trademark Protection*, 40 HARV. L. REV. 813, 832 (1927).

188. 15 U.S.C. § 1125(c)(2)(A).

marks receive the almost “in gross” rights provided by antidilution laws.¹⁸⁹ This provision was intended to prevent the use of federal dilution law to protect marks that have only achieved so-called “niche” fame (that is, fame in only a specialized market). However, recent case law has threatened this limit on dilution by finding marks famous that may not rise to the level of fame intended by the statute.¹⁹⁰

Fame is a concept that is closely related to strength of the mark. Strength of the mark, however, denotes the distinctiveness of a mark. A mark may be famous and not distinctive.¹⁹¹ The TDRA includes the additional requirement that a mark must be distinctive as well as famous.¹⁹² Whether a mark is famous is determined by examining a list of non-exclusive factors under the TDRA.¹⁹³ Many of these factors are similar to those examined in an analysis of the strength of a mark for purposes of trademark infringement.¹⁹⁴

The extent of unauthorized third party usage of the mark is highly relevant to a fame analysis.¹⁹⁵ Substantial third party use of a particular mark can impact whether a mark is famous or reduce the fame of the mark.¹⁹⁶ Thus, mark holders

189. See Bd. of Regents, Univ. of Tex. Sys. v. KST Elec., Ltd., 550 F. Supp. 2d 657, 674 (W.D. Tex. 2008) (noting that famous marks under the TDRA are “household names”).

190. Dall. Cowboys Football Club, Ltd. v. America’s Team Props., Inc., 616 F. Supp. 2d 622, 636 (N.D. Tex. 2009) (finding the “America’s Team” mark famous amongst the general consuming public); Univ. of Kan. v. Sinks, 644 F. Supp. 2d 1287, 1306-07 (D. Kan. 2008) (finding numerous marks concerning the University of Kansas famous).

191. V Secret Catalogue, Inc. v. Moseley, 558 F. Supp. 2d 734, 746 (W.D. Ky. 2008).

192. 15 U.S.C. § 1125(c). Since the TDRA requires distinctiveness as well as fame, the cases concerning third party use and secondary meaning are relevant to the distinctiveness prong of a federal dilution action.

193. Those factors include:

- (i) The duration, extent, and geographic reach of advertising and publicity of the mark, whether advertised or publicized by the owner or third parties.
- (ii) The amount, volume, and geographic extent of sales of goods or services offered under the mark.
- (iii) The extent of actual recognition of the mark.
- (iv) Whether the mark was registered . . . on the principal register.

Id. § 1125(c)(2)(A).

194. *Id.*

195. Fame can be viewed as a higher level of secondary meaning than what is required to establish acquired distinctiveness for a trademark. However, a mark can be famous and not distinctive. Moreover, although the TDRA removed “third party” use from the statute in determining fame, third party use must still be relevant because it affects the strength of the mark. See 4 MCCARTHY, *supra* note 173, at § 24:106 (“The lack of an explicit mention of third party use in the ‘fame’ factors of the TDRA cannot, in the author’s view, mean that third party use is not a relevant factor to consider in the determination of whether a mark is deserving of the high status of being called ‘famous’ for purposes of dilution protection. Evaluating mark strength and fame by the degree of third party use is a method often used in traditional infringement analysis.”).

196. *Id.* § 24:106; *id.* § 24:108; see Univ. of Kan. v. Sinks, 565 F. Supp. 2d 1216, 1231-33

are incentivized to enforce their trademark rights vigorously to prohibit third party usage of the mark (regardless of what product or service is used), particularly since niche fame is no longer allowed under the TDRA.¹⁹⁷ This problem is further exacerbated by the fact that fame is determined from the perspective of consumers.

There are generally two types of dilution: blurring and tarnishment. Third party use may also impact whether a mark is subject to dilution by blurring—a mark’s ability to serve as an indicator of a sole source, and the particular third party use of a mark may result in tarnishment—the association of the mark with goods of poor quality or some sexual reference. The TDRA includes factors to determine whether dilution by blurring is likely.¹⁹⁸ For example, the TDRA specifically provides that the inherent and acquired distinctiveness of the mark is relevant along with the degree of recognition of the mark.¹⁹⁹ Courts may analyze these factors in much the same way as they analyze strength of the mark in the trademark infringement analysis. The TDRA also provides that a relevant factor in determining blurring is whether the owner of the famous mark has engaged in substantially exclusive use of the mark.²⁰⁰ This factor specifically

(D. Kan. 2008) (admitting evidence of third party use of a mark for purposes of proving strength and fame of the mark); *Milbank Tweed Hadley & McCloy, LLP v. Milbank Holding Corp.*, No. CV-06-187-RGK, 2007 WL 1438114, at *5 (C.D. Cal. Feb. 23, 2007) (“[E]vidence demonstrating widespread third-party use of the word ‘Milbank’ is relevant because, when a mark is in widespread use, it may not be famous for the goods or services of one business.”); *Smith v. Wal-Mart Stores, Inc.*, 475 F. Supp. 2d 1318, 1323 (N.D. Ga. 2007) (noting that third-party use of mark is relevant to the fame and distinctiveness inquiry for a tarnishment action).

197. See 15 U.S.C. § 1125(c)(1).

198. *Id.* § 1125(c)(2)(B). That section, in relevant part, provides:

For purposes of paragraph (1), “dilution by blurring” is association arising from the similarity between a mark or trade name and a famous mark that impairs the distinctiveness of the famous mark. In determining whether a mark or trade name is likely to cause dilution by blurring, the court may consider all relevant factors, including the following:

- (i) The degree of similarity between the mark or trade name and the famous mark.
- (ii) The degree of inherent or acquired distinctiveness of the famous mark.
- (iii) The extent to which the owner of the famous mark is engaging in substantially exclusive use of the mark.
- (iv) The degree of recognition of the famous mark.
- (v) Whether the user of the mark or trade name intended to create an association with the famous mark.
- (vi) Any actual association between the mark or trade name and the famous mark.

Id.

199. *Id.* §§ 1125(c)(2)(B)(ii), (iv).

200. *Id.* § 1125(c)(2)(B)(iii).

takes into account third party use.²⁰¹ This factor alone provides a substantial incentive for trademark owners to police third party usage—even in markets that are unrelated.²⁰² Accordingly, the availability of a cause of action for dilution by blurring provides strong incentives for trademark owners to police their marks vigorously because of the combination of strength of the mark measured by distinctiveness and fame and substantial exclusive use of the mark by the owner.²⁰³ The TDRA also provides several exclusions to trademark dilution actions designed to protect free speech and other interests;²⁰⁴ however, it is unclear at this time whether those protections will adequately protect free speech and other interests and deter policing efforts by trademark holders.²⁰⁵

201. See *Miss Universe, L.P. v. Villegas*, 672 F. Supp. 2d 575, 594 (S.D.N.Y. 2009) (“The plaintiff opposes registration and use of marks that it perceives infringes to infringe its own, and it has succeeded in obtaining injunctions against the use of marks like ‘Miss Nude USA’ and ‘Mrs. Nude USA,’ and ‘Mrs. USA,’ ‘Mrs. Universe,’ and ‘Mrs. [State or Locality] USA.’ It claims to spend more than three hundred thousand dollars each year to enforce its trademarks. ‘Miss USA,’ in short, is not a trademark whose value is already substantially diluted by its owner’s carelessness or indifference in protecting it. Miss Universe takes care to guard both the mark’s value and its exclusive use of that mark.” (internal citations omitted)); *V Secret Catalogue, Inc. v. Moseley*, 558 F. Supp. 2d 734, 746 (W.D. Ky. 2008) (“V Secret asserts that it is engaged in substantially exclusive use of the VICTORIA’S SECRET mark. In support of this contention, Vice President Kriss avers that V Secret actively and aggressively polices its mark against unauthorized and dilutive uses by others. It pursues unauthorized uses through cease and desist letters, trademark oppositions, and litigation. . . . We therefore find the requisite substantially exclusive use of the mark.” (internal citation omitted)).

202. In examining the second and third factors, Professor Barton Beebe notes that a mark such as apple for computers is “arbitrary” and thus inherently distinctive. Nevertheless, it is used by several other firms, most notably the record production company and the bank that go by the same name. This may function to narrow the scope of anti-blurring protection for such marks even though they are inherently distinctive. This is a progressive development and makes sense from the point of view of competition. Firms do not likely need to use the entirely fanciful marks . . . of other firms, but they may need or want to use the marks of other firms that are arbitrary . . . with respect to those other firms’ products.

Barton Beebe, *A Defense of the New Federal Trademark Antidilution Law*, 16 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 1143, 1171 (2006).

203. Mark owners are also incentivized to police “tarnishing” uses of their marks because they are unlikely to want their marks associated with low quality goods or sexual connotations.

204. 15 U.S.C. § 1125(c)(3).

205. Cf. *Starbucks Corp. v. Wolfe’s Borough Coffee, Inc.*, 588 F.3d 97, 112 (2d Cir. 2009) (“As evident from the statutory language, Black Bear’s use of the Charbucks Marks cannot qualify under the parody exception because the Charbucks Marks are used ‘as a designation of source for [Black Bear’s] own goods[, i.e., the Charbucks line of coffee].’”); *Louis Vuitton Malletier S.A. v. Haute Diggity Dog, LLC*, 507 F.3d 252, 266 (4th Cir. 2007) (“We begin by noting that parody is not automatically a complete *defense* to a claim of dilution by blurring where the defendant uses the parody as its own designation of source, i.e., *as a trademark*.” (emphases in original)). In an

5. *Fair Use.*—Under the classic fair use defense to trademark infringement, an alleged infringer is permitted use of a protected trademark in a descriptive way.²⁰⁶ Potential infringers can only take advantage of the defense if they used the mark in good faith. Despite the existence of this defense, mark holders are incentivized to police their marks and bring actions against third parties because of the reasons discussed in this section. They are also incentivized to threaten an action which may cause the alleged infringer to stop using the supposedly diluting or infringing mark. Moreover, if a mark holder sends a cease and desist letter to a potential infringer, it may result in the removal of good faith and thus

empirical study, Professor Clarisa Long found that “[j]udicial enforcement of dilution law is not robust today and has been eroding over time.” Clarisa Long, *Dilution*, 106 COLUM. L. REV. 1029, 1031 (2006). Professor Long does note that “[i]t could well be the case that dilution law is a powerful bargaining chip in cease-and-desist letters and in negotiations entirely outside the litigatory arena.” *Id.* In another empirical study analyzing cases reported one year after passage of the TDRA, Professor Beebe found that “[i]n the few reported opinions that addressed a speech-related issue, the TDRA’s new, more robust ‘[e]xclusions’ from antidilution protection played no appreciable role.” Barton Beebe, *The Continuing Debacle of U.S. Antidilution Law: Evidence from the First Year of Trademark Dilution Revision Act Case Law*, 24 SANTA CLARA COMPUTER & HIGH TECH. L.J. 449, 467 (2008) [hereinafter Beebe, *Antidilution Law*] (citations omitted). The TDRA also provides that an allegedly diluting use must be a use in commerce. *See* 15 U.S.C. § 1125(c)(1). This may provide some protection for free expression.

206. There is a common law version of the defense as well as a federal codification of the common law defense. The following is the common law defense of classical fair use as stated in the *Restatement of Unfair Competition*:

In an action for infringement of a trademark, trade name, collective mark, or certification mark, it is a defense that the term used by the actor is descriptive or geographically descriptive of the actor’s goods, services, or business, or is the personal name of the actor or a person connected with the actor, and the actor has used the term fairly and in good faith solely to describe the actor’s goods, services, or business or to indicate a connection with the named person.

RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 28 (1995). The statutory fair use defense provides:

That the use of the name, term, or device charged to be an infringement is a use, otherwise than as a mark, of the party’s individual name in his own business, or of the individual name of anyone in privity with such party, or of a term or device which is descriptive of and used fairly and in good faith only to describe the goods or services of such party, or their geographic origin

15 U.S.C. § 1115(b)(4).

The statutory fair use defense includes three elements: the alleged infringing mark was (1) not used as a trademark; (2) used fairly and in good faith; and (3) used only to describe its own goods or services. *See* ETW Corp. v. Jireh Publ’g, Inc., 332 F.3d 915, 920 (6th Cir. 2003); Packman v. Chi. Trib. Co., 267 F.3d 628, 639 (7th Cir. 2001); *see also* KP Permanent Make-Up, Inc. v. Lasting Impression I, Inc., 543 U.S. 111, 117-19 (2004) (discussing the statutory fair use defense).

the option for the potential infringer to take advantage of the defense.²⁰⁷ This provides an incentive to mark holders to send cease and desist letters, which may also result in third parties ceasing use of the mark.

6. *Abandonment*.—In addition to other evidence, failure to enforce a mark against third parties may result in the loss of the mark through abandonment.²⁰⁸ Language concerning trademark abandonment is used by numerous courts in different contexts.²⁰⁹ Generally, trademark abandonment in the federal infringement context occurs in two situations. The first involves nonuse of the mark with intent not to resume use.²¹⁰ The second involves an act or omission that may lead to a mark essentially becoming generic.²¹¹ Evidence of failure to

207. Cf. *Elvis Presley Enters., Inc. v. Capece*, 141 F.3d 188, 205 (5th Cir. 1998) (“Any acts after receiving a cease and desist letter are at the defendant’s own risk because it is on notice of the plaintiff’s objection to such acts.”); *Bd. of Regents, Univ. of Tex. Sys. v. KST Elec., Ltd.*, 550 F. Supp. 2d 657, 665 (W.D. Tex. 2008) (noting the same); *Frito-Lay, Inc. v. Bachman Co.*, 704 F. Supp. 432, 437 (S.D.N.Y. 1989) (“Awareness of pre-existing use of a mark can give rise to an inference of bad faith.”). But see *Packman*, 267 F.3d at 642 (finding no bad faith based only on the alleged infringer’s knowledge of the existence of owner’s mark); *Streetwise Maps, Inc. v. VanDam, Inc.*, 159 F.3d 739, 746 (2d Cir. 1998) (“Defendants’ failure to perform an official trademark search, even with the knowledge that plaintiff marketed its maps under the Streetwise name, does not standing alone prove that they acted in bad faith.”); *Int’l Stamp Art, Inc. v. U.S. Postal Serv.*, No. 1:02-CV-2459-TWT, 2005 WL 3947951, at *5 (N.D. Ga. May 27, 2005) (finding the alleged infringer’s knowledge of a mark insufficient for a finding of bad faith), *aff’d*, 456 F.3d 1270 (11th Cir. 2006); *Wonder Labs, Inc. v. Proctor & Gamble Co.*, 728 F. Supp. 1058, 1064 (S.D.N.Y. 1990) (refusing to find bad faith where plaintiff’s primary proof of bad faith was the alleged infringer’s failure to stop using the mark after receiving a cease and desist letter).

208. See *Glover v. Ampak, Inc.*, 74 F.3d 57, 60 (4th Cir. 1996) (“[E]vidence of others’ use of the . . . [purported trademark] might be probative of trademark dilution or, if the usage was known by . . . [the alleged trademark owner], abandonment.”); *Quality Inns Int’l, Inc. v. McDonald’s Corp.*, 695 F. Supp. 198, 214 (D. Md. 1988) (“Third-party uses permitted by the owner of a mark may also be probative of the abandonment of a mark by the owner.”).

209. See 3 MCCARTHY, *supra* note 173, at § 17:17 (disagreeing with courts using language concerning abandonment in the context of a mark holder failing to enforce its mark against third parties). Professor McCarthy would categorize failure to enforce a mark not as abandonment, but as impacting the strength of the mark. *Id.*

210. Section 1127 of 15 U.S.C., concerning “abandonment,” provides:

A mark shall be deemed to be “abandoned” if . . . its use has been discontinued with intent not to resume such use. Intent not to resume may be inferred from circumstances. Nonuse for 3 consecutive years shall be *prima facie* evidence of abandonment. “Use” of a mark means the bona fide use of such mark made in the ordinary course of trade, and not made merely to reserve a right in a mark.

15 U.S.C. § 1127 (2006).

211. Section 1127 of 15 U.S.C., in relevant part, states:

A mark shall be deemed to be “abandoned” if . . . any course of conduct of the owner, including acts of omission as well as commission, causes the mark to become the generic name for the goods or services on or in connection with which it is used or

enforce a mark is likely particularly relevant in the latter situation; however, that same evidence may be relevant in the former situation to demonstrate an intent not to resume use of the mark. Thus, trademark holders have an incentive to enforce their rights or lose them by abandonment.²¹² “Once a mark has been held abandoned, it is free for all to use and falls into the public domain.”²¹³

7. *Equitable Defenses.*—Failure to police third party usage and other conduct may also result in the loss of trademark rights through either laches or some other equitable defense such as estoppel or perhaps even acquiescence.²¹⁴ The application of these defenses results in a party essentially receiving an

otherwise to lose its significance as a mark. Purchaser motivation shall not be a test for determining abandonment under this paragraph.

15 U.S.C. § 1127; *see also* Sung In, Note, *Death of a Trademark: Genericide in the Digital Age*, 21 REV. LITIG. 159, 173 (2002) (“[A] reasonably diligent mark owner would invariably police . . . [improper generic use on a competitor’s similar product] as it falls within the ambit of trademark infringement, and failure to do so could end mark ownership.”).

212. Herman Miller, Inc. v. Palazzetti Imports & Exports, Inc., 270 F.3d 298, 317 (6th Cir. 2001) (“Although it appears unlikely that failure to prosecute, by itself, can establish that trade dress has been abandoned, it is possible that, in extreme circumstances, failure to prosecute may cause trade dress rights to be extinguished by causing a mark to lose its significance as an indication of source.”); Sweetheart Plastics, Inc. v. Detroit Forming, Inc., 743 F.2d 1039, 1048 (4th Cir. 1984) (“A party may be deemed to have abandoned a valid trademark by permitting such excessive adverse use of a mark that it has lost its significance as an indication of origin.”); *see also* Pebble Beach Co. v. Tour 18 I Ltd., 155 F.3d 526, 542-43 (5th Cir. 1998) (finding that aggressive policing of a mark is relevant to the issue of abandonment), *abrogated by* Amazing Spaces, Inc. v. Metro Mini Storage, 608 F.3d 225 (5th Cir. 2010).

213. *Quality Inns Int’l, Inc.*, 695 F. Supp. at 214. However, it is difficult for a mark to be held to be abandoned even with extensive third party use. *See Sweetheart Plastics*, 743 F.2d at 1047-48; *Wallpaper Mfrs. Ltd. v. Crown Wallcovering Corp.*, 680 F.2d 755, 766 (C.C.P.A. 1982) (“[A]n owner is not required to act immediately against every possibly infringing use to avoid a holding of abandonment. Such a requirement would unnecessarily clutter the courts.” (internal citation omitted)). In *Board of Governors of University of North Carolina v. Helpingsstine*, 714 F. Supp. 167 (M.D.N.C. 1989), the court found that notwithstanding many years of third party use, the University of North Carolina had not abandoned its marks by failing to enforce them against third parties. *Id.* at 171-72. The court determined that the university had continued to use its marks and that those marks had not lost their significance as indicators of source—thus, the marks were not abandoned. *Id.* If there is residual goodwill, some courts may not find abandonment. *See Michael S. Denniston, Residual Good Will in Unused Marks—The Case Against Abandonment*, 90 TRADEMARK REP. 615, 634 (2000) (“If good will in the mark persists—‘residual good will’—courts should protect the public interest by refusing to allow a junior user to appropriate the mark.”); McKenna, *Normative Foundations*, *supra* note 15, at 1893 n.233 (citing cases).

214. *See JAY DRATLER, JR., LICENSING OF INTELLECTUAL PROPERTY* § 3.05 (1999) (discussing equitable principles that are effectively “implied licenses”). In *American International Group, Inc. v. American International Bank*, 926 F.2d 829 (9th Cir. 1991), the court denied injunctive relief because of laches in a trademark matter.

implied license to use a mark if a mark holder has failed to enforce it.²¹⁵ For example, laches may be an effective defense against enforcement of trademark rights against a particular alleged infringer. A trademark holder may be subject to a laches defense when it unreasonably or negligently delays enforcing its rights.²¹⁶ Thus, laches may be available as a defense where an alleged infringer can demonstrate: “(1) lack of diligence by the party against whom the defense is asserted, and (2) prejudice to the party asserting it.”²¹⁷ While this defense may only be available to a particular accused infringer, it nevertheless gives mark holders yet another incentive to police their rights.²¹⁸ If a mark holder is not adequately policing its mark, it may lose the right to stop an alleged infringer. This is particularly true if a trademark owner does not acquire federal trademark registration and nationwide constructive use rights to protect areas in which the trademark holder would like to expand.²¹⁹ Acquiescence may similarly estop a trademark holder from successfully enjoining an alleged infringer, but it also

215. Cf. DRATLER, *supra* note 214, at § 3.04(8)(d) (stating that it is unnecessary for the court to find an implied license because of the defense of fair use; also, absent the ability for the court to imply provisions concerning quality control, it may be difficult for a court to impose those terms on a purported trademark licensee without evidence of an “existing contract or a longstanding business relationship”).

216. *E.g.*, Nartron Corp. v. STMicroelecs., Inc., 305 F.3d 397, 408 (6th Cir. 2002). Laches may be difficult to prove because a mere long delay in enforcement of a mark is insufficient to find the defense applicable. *See* Carl Zeiss Stiftung v. VEB Carl Zeiss Jena, 433 F.2d 686, 704 (2d Cir. 1970) (finding a fourteen-year delay in bringing suit insufficient for the laches defense).

217. *Nartron Corp.*, 305 F.3d at 408; *see also* Miller v. Glenn Miller Prods., Inc., 454 F.3d 975, 997, 999 (9th Cir. 2006) (“In considering whether a plaintiff’s delay was unreasonable, courts consider: (1) the length of the delay, measured from the time the plaintiff knew or should have known about his potential cause of action, and (2) whether the plaintiff’s delay was reasonable, including whether the plaintiff has proffered a legitimate excuse for his delay. . . . A defendant may establish prejudice by showing that during the delay, it invested money to expand its business or entered into business transactions based on his presumed rights.”).

218. Some jurisdictions may recognize the doctrine of progressive encroachment, which will excuse delay in bringing suit until a competitor has used a trademark in sufficient manner as to constitute a likelihood of confusion. *See* AM Gen. Corp. v. DaimlerChrysler Corp., 311 F.3d 796, 823-24 (7th Cir. 2002); Kellogg Co. v. Exxon Corp., 209 F.3d 562, 570-73 (6th Cir. 2000).

219. Professor Landes and Judge Posner explain that

[t]he doctrine of laches . . . forces *A*[, the prior user,] to internalize *B*’s[, the subsequent user’s,] cost of duplication when appropriate. If *A* has reason to know that *B* is proceeding to develop a similar mark in ignorance of *A*’s prior use, *A*, unless he has registered his mark, must, on pain of not being able to use it in his expansion markets, warn *B* off. The cost to *A* of preventing the collision of the two marks is less than that of *B*. But if *B*, rather than proceeding in ignorance of *A*’s prior use (that is, in good faith), has deliberately copied *A*’s mark, the costs of duplication are self-imposed; *B* is the cheapest cost avoider and the defense of laches to *A*’s suit for infringement is rejected.

LANDES & POSNER, ECONOMIC STRUCTURE, *supra* note 10, at 183.

only applies to the alleged infringer.²²⁰ Thus, equitable defenses may provide an additional incentive for firms to police and enforce their marks.

8. *Licensing*.—Generally, in the licensing context, if a licensor fails to exercise control over the quality of the goods and services used in connection with a licensed mark, the mark holder may lose its trademark rights.²²¹ This requirement thus provides an incentive for trademark holders to police potential licensees.²²² This makes sense from a consumer search cost perspective since a use of the mark by licensees that is inconsistent with the use by the mark holder may result in consumer confusion.²²³

9. *Other Reasons to Police*.—There are several other reasons for mark holders to police usage of their marks. Firms may attempt to acquire a competitive advantage by foreclosing a competitor's use of a word or symbol necessary or effective for purposes of competition. A trademark owner may send a cease and desist letter to an alleged infringer, which may result in that party stopping use of the mark, even if the alleged infringer may have a right to use that mark to avoid costly and uncertain litigation. Moreover, trademark owners who do not police their marks risk losing market share.²²⁴ Additionally, trademark owners may send cease and desist letters to improve their chances of receiving an award of attorney fees.²²⁵ Receipt of a cease and desist letter followed by continued use by the accused infringer (along with other facts) may result in a finding of an exceptional case leading to an award of attorney fees.²²⁶

220. *Quality Inns Int'l, Inc. v. McDonald's Corp.*, 695 F. Supp. 198, 213 (D. Md. 1988). *But see DRATLER, supra* note 214, § 3.05(1)(a) (stating that estoppel may work in the opposite direction if a trademark owner sends a cease and desist letter and then fails to follow up with an infringement action).

221. *Miller*, 454 F.3d at 993; RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 33 (1995) (“Failure of the licensor to exercise reasonable control over the use of the designation by the licensee can result in abandonment . . .”).

222. Once a licensee exceeds the scope of the license, the licensor can bring a cause of action for trademark infringement. *See Miller*, 454 F.3d at 996 (recognizing that the relationship between scope of license and trademark infringement action remains).

223. Notably, courts have increasingly relaxed the requirement of quality control, recognizing that necessary quality control may be based upon a long-term or other special relationship. *See TMT N. Am., Inc. v. Magic Touch GmbH*, 124 F.3d 876, 885-86 (7th Cir. 1997); *Ky. Fried Chicken Corp. v. Diversified Packaging Corp.*, 549 F.2d 368, 387 (5th Cir. 1977); *see also* RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 33 cmt. c (1995). At least one commentator argues that the quality control requirement for licensing should be abandoned. *See generally* Kevin Parks, “*Naked*” Is Not a Four-Letter Word: Debunking the Myth of the “Quality Control Requirement” in Trademark Licensing, 82 TRADEMARK REP. 531 (1992).

224. Michael J. Freno, *Trademark Valuation: Preserving Brand Equity*, 97 TRADEMARK REP. 1055, 1063 (2007) (“[A]n infringer’s unauthorized use of a trademark is likely cutting into the trademark owner’s sales and profits, which directly decreases the brand value under an expected income valuation.” (internal citation omitted)).

225. 15 U.S.C. § 1117 (2006 & Supp. 2009).

226. Section 1117 of 15 U.S.C., in relevant part, provides: “The court in exceptional cases

10. Conclusion.—The structure of trademark law provides a strong incentive for trademark owners to police their marks vigorously because of the threat of a loss or lessening of rights—even against non-competitors—or the lure of a competitive advantage. As Judge Posner stated:

[IHSA] could have sued . . . [CBS] . . . for using its trademark [March Madness] to promote CBS's broadcast of the NCAA championship. And it could have supplicated them not to spoil its trademark by using it to name something else. A serious trademark holder is assiduous in endeavoring to convince dictionary editors, magazine and newspaper editors, journalists and columnists, judges, and other lexicographically influential persons to avoid using his trademark to denote anything other than the trademarked good or service. These efforts sometimes succeed. IHSA was not assiduous.²²⁷

may award reasonable attorney fees to the prevailing party.” 15 U.S.C. § 1117(a); *see also* Tap Publ’ns, Inc. v. Chinese Yellow Pages (N.Y.) Inc., 925 F. Supp. 212, 224 (S.D.N.Y. 1996) (“In order to find that a case is ‘exceptional’ under this provision, there must be a showing of bad faith.”). In *International Star Class Yacht Racing Ass’n v. Tommy Hilfiger, U.S.A., Inc.*, 80 F. 3d 749 (2d Cir. 1996), the court found that the district court made several clearly erroneous findings on bad faith and willful infringement and failed to consider all of the evidence concerning those two findings. *Id.* at 753-55. The court noted that

[t]he suit gave Hilfiger notice of its potential trademark violation, and . . . Hilfiger nonetheless continued to sell its merchandise with the infringing mark, racking up over \$3 million in sales, without regard for the rights of ISCYRA. As counsel for Hilfiger admitted during oral argument, Hilfiger was betting on the fact that ISCYRA would not prevail in its suit. Hilfiger lost that bet, and should not escape the consequences of its conduct.

Id. at 754. Notably, on remand, the district court did not find bad faith or willful infringement. *Int’l Star Class Yacht Racing Ass’n v. Tommy Hilfiger, U.S.A., Inc.*, No. 94 CIV 2663 (RPP), 1999 WL 108739, at *3 (S.D.N.Y. Mar. 3, 1999). The district court explained:

In view of the fact that (1) Hilfiger was only using the Star Class mark as an embellishment on its Nantucket line of clothing, which was exceedingly well trademarked with the Hilfiger name designating the source of the goods, (2) there is no showing that Hilfiger intended to copy a trademark, (3) Hilfiger’s conduct was consistent with the advice of its attorneys concerning when a full search would be required, and (4) Hilfiger was reasonably advised by its attorneys that the Star Class mark was weak, the Court finds that the evidence is insufficient to support a finding of bad faith.

Id. at *3.

227. *Ill. High Sch. Ass’n v. GTE Vantage Inc.*, 99 F. 3d 244, 246 (7th Cir. 1996). In *GTE Vantage Inc.*, the Seventh Circuit did not allow the senior user to prohibit use of its mark after the senior user failed to adequately police its rights, and the public appropriated the mark. *Id.* at 246-48; *see* Joseph E. Washington, *The Impact of Public and Media Use on Trademark Rights: An Analysis of Illinois High School Ass’n v. GTE Vantage Inc. and “Dual-Use” Terms*, 48 CATH. U. L. REV. 605, 609 (1999) (“[T]he Seventh Circuit allowed both entities to continue using the mark.

Notably, the drive to enforce trademarks and the difficulty involved in actively policing them has led to the development of many businesses that offer trademark monitoring services, including watching for trademark filings that may be infringing and policing the use of trademarks on the Internet and other uses.²²⁸ Trademark owners who fail to police their marks take great risks with the value of their marks under trademark law.

C. The Expansion of Trademark Law and the Inadequacy of the Consumer Search Cost Theory

The failure of the search cost theory to provide adequate limits to trademark law has led to the adoption and development of new doctrines that are criticized as unproductively expansive. In some cases, protection is afforded where it should not be provided, and trademark law is offering remedies for harms that it should not be concerned about preventing. Some specific areas where search cost theory—and thus, consumer perception—has failed to provide sufficient limits to trademark rights include merchandising rights, dilution, new theories of confusion, the trademark “use” requirement for infringement, the *Dawn Donut* rule, and maintenance of the public domain.

1. *Merchandising*.—In merchandising rights cases, courts have allowed protection for marks used in connection with goods and services that are ancillary to the primary good or service with which the mark is used.²²⁹ In those cases,

... The decision suggests, therefore, that trademark owners should be diligent in policing the use of their trademarks by others or risk losing protection of their trademark rights.” (internal citations omitted)). The court in *GTE Vantage Inc.* also noted that a senior user cannot take a mark that the public is using and render the public speechless. *GTE Vantage Inc.*, 99 F.3d at 247. This decision seems to indicate that a mark holder should not allow public and media usage and that the mark holder should immediately assert its rights before any reliance interest by the public evolves.

228. See, e.g., *AdGooroo Products Overview*, ADGOOROO SEM INSIGHT, http://www.adgooroo.com/adgooroo_products_overview.php (last visited Mar. 25, 2011); BRAND PROTECTION AGENCY, <http://www.brandprotectionagency.com/> (last visited Mar. 25, 2011); BRANDVERITY, <https://www.brandverity.com/> (last visited Mar. 25, 2011); CORP. SERV. CO., https://www.cscglobal.com/global/web/csc/brand-monitoring.html?iq_id=8779138 (last visited Mar. 25, 2011); *Trademark Infringement Monitoring Service*, CYBERALERT, http://www.cyberalert.com/appTrademark_infringement.html (last visited Mar. 25, 2011); PPC ENFORCER, <http://ppcenforcer.com> (last visited Mar. 25, 2011); THOMSON COMPUTUMARK, <http://compumark.thomson.com/> (last visited Mar. 25, 2011); TRADEMARK CONFIDENCE, <http://www.trademarkconfidence.com/welcome.html> (last visited Mar. 25, 2011); see also CHILLING EFFECTS CLEARINGHOUSE, <http://www.chillingeffects.org> (last visited Mar. 25, 2011).

229. Bd. of Supervisors for La. State Univ. Agric. & Mech. Coll. v. Smack Apparel Co., 550 F.3d 465, 483-84 (5th Cir. 2008) (“Smack’s use of the Universities’ colors and indicia . . . [on t-shirts] is designed to create the illusion of affiliation with the Universities and essentially obtain a ‘free ride’ by profiting from confusion among the fans of the Universities’ football teams who desire to show support for and affiliation with those teams.”); Univ. of Ga. Athletic Ass’n v. Laite, 756 F.2d 1535, 1546 n.28 (11th Cir. 1985) (“The record in the instant case reveals that, in one

there is very little chance that such merchandising uses will raise consumer search costs for the primary good or service with which the mark is used in connection, unless one accepts that possible tarnishment increases consumer search costs. Indeed, scholars have been critical of the expansion of trademark protection to merchandising.²³⁰ Notably, the recognition of a mark in connection with certain goods or services may increase if other companies produce shirts or

week, at least ten to fifteen members of the public contacted UGAA to inquire about the connection between ‘Battlin’ Bulldog Beer’ and the University of Georgia. This evidence indicates that . . . at least some members of the public *do* assume that products bearing the mark of a school or a sports team are sponsored or licensed by the school or team.” (emphasis in original) (internal citation omitted)); *Auburn Univ. v. Moody*, No. 3:08cv796-CSC, 2008 WL 4877542, at *8 (M.D. Ala. Nov. 4, 2008) (“In . . . [other] words, confusion can occur when it appears that the owner of the marks, in this case, Auburn University, sponsored Moody’s use of its marks.”); *cf. Au-Tomotive Gold, Inc. v. Volkswagen of Am., Inc.*, 457 F.3d 1062, 1074 (9th Cir. 2006) (stating that trademarks were infringed by the defendant’s production of license plates, license plate covers, and key rings using the plaintiff’s marks); *Nat’l Bus. Forms & Printing, Inc. v. Ford Motor Co.*, No. H-08-1906, 2009 WL 3570387, at *5 (S.D. Tex. Oct. 30, 2009) (finding that NBFP “used” Ford’s marks in connection with goods because the decals and stickers were the goods). The case that contributed to the expansion of merchandising rights is *Boston Professional Hockey Ass’n, Inc. v. Dallas Cap & Emblem Manufacturing, Inc.*, 510 F.2d 1004 (5th Cir. 1975). In *Boston Professional Hockey Ass’n*, the court held that the defendant infringed the plaintiff’s trademark by reproducing the mark as a patch, knowing that purchasers would know that the copies were the plaintiff’s mark. *Id.* at 1011 (“When defendant causes plaintiffs’ marks to be embroidered upon emblems which it later markets, defendant uses those marks in connection with the sale of goods as surely as if defendant had embroidered the marks upon knit caps.”). *Boston Professional Hockey Ass’n* has been limited in a subsequent Fifth Circuit case, and the Ninth Circuit declined to follow it. *See Supreme Assembly, Order of Rainbow for Girls v. J.H. Ray Jewelry Co.*, 676 F.2d 1079, 1085 (5th Cir. 1982); *Int’l Order of Job’s Daughters v. Lindeburg & Co.*, 633 F.2d 912, 918 (9th Cir. 1980) (“It is not uncommon for a name or emblem that serves in one context as a collective mark or trademark also to be merchandised for its own intrinsic utility to consumers. We commonly identify ourselves by displaying emblems expressing allegiances. Our jewelry, clothing, and cars are emblazoned with inscriptions showing the organizations we belong to, the schools we attend, the landmarks we have visited, the sports teams we support, the beverages we imbibe. Although these inscriptions frequently include names and emblems that are also used as collective marks or trademarks, it would be naive to conclude that the name or emblem is desired because consumers believe that the product somehow originated with or was sponsored by the organization the name or emblem signifies.”); *Ky. Fried Chicken Corp. v. Diversified Packaging Corp.*, 549 F.2d 368, 389 (5th Cir. 1977) (“Under these circumstances we do not believe *Boston Hockey* equates knowledge of the *symbol*’s source with confusion sufficient to establish trademark infringement” (emphasis in original)).

230. See Dogan & Lemley, *Merchandising Right*, *supra* note 67, at 465 (“When a trademark is sold, not as a source indicator, but as a desirable feature of a product, competition suffers—and consumers pay—if other sellers are shut out of the market for that feature.”). But see generally Robert C. Denicola, *Institutional Publicity Rights: An Analysis of the Merchandising of Famous Trade Symbols*, 62 N.C. L. REV. 603 (1984) (proposing a merchandising right for trademarks).

other items using the mark, and thus may lower search costs.

Trademark owners do have some valid concerns with such usage.²³¹ Because the consumer confusion analysis is driven by what the consumer may perceive, merchandising may also lead to some form of sponsorship or association confusion, where the consumer believes that the merchandise comes from the mark holder or is “officially licensed.”²³² This is another example of how a trademark owner may control the consumer’s perception of a mark and thereby expand its trademark rights.²³³ A trademark owner may be concerned with the quality of the merchandise used in connection with its mark.²³⁴ A junior user of a mark may use the mark owner’s mark on merchandise of a dissimilar quality than the mark owner’s merchandise. That could lead to tarnishment and arguably some type of confusion actionable under a consumer search cost theory. Moreover, the value of licensing trademarks as merchandise will continue to drive trademark owners to enforce licensing uses, and the consumer search cost theory allows them to do so. Indeed, one source states that in 2009, the estimated worldwide retail sales of licensed products amounted to \$182.4 billion.²³⁵

231. There are additional reasons why trademark doctrine provides an incentive to police merchandising uses. For example, a trademark owner may worry that third party usage in the merchandising context will lead to a weakening of the strength of the mark in its primary context or create an argument for abandonment. *See supra* notes 182-86 and accompanying text.

232. *Cf. Univ. of Ga. Athletic Ass’n*, 756 F.2d at 1546 n.28 (“Furthermore, in our view, most consumers who purchase products containing the name or emblem of their favorite school or sports team would prefer an officially sponsored or licensed product to an identical non-licensed product. Were this not true, the word ‘official’ would not appear in so many advertisements for such products.”).

233. *See* 4 MCCARTHY, *supra* note 173, at § 24:9 (“If consumers think that most uses of a trademark require authorization, then in fact they will require authorization because the owner can enjoin consumer confusion caused by unpermitted uses or charge for licenses. And if owners can sue to stop unauthorized uses, then only authorized uses will be seen by consumers, creating or reinforcing their perception that authorization is necessary.”); Gibson, *supra* note 107, at 907 (noting how trademark licensing practices create a feedback loop leading to expansive trademark protection); Lemley, *supra* note 2, at 1708 (“It is possible that consumers have come to expect that ‘Dallas Cowboys’ caps are licensed by the Cowboys, not because they serve a trademark function, but simply because the law has recently required such a relationship. If this expectation exists, consumers may be confused if the law changes.”); Lunney, *supra* note 1, at 397 (“Whether such use is likely to generate confusion, of sponsorship or otherwise, is thus circular—both a reflection and the determinant of the trademark owner’s rights.”); McKenna, *Trademark Use*, *supra* note 14, at 821-22 (noting the feedback loop).

234. 4 MCCARTHY, *supra* note 173, at § 24:8 (noting that “[c]onsumers will look to the university to be the party responsible for poor quality goods, even though they know full well that the university is not a manufacturer”). Professor McCarthy also states that in his opinion, “the ordinary consumer seeing non-licensed university-marked wearing apparel is likely to be confused as to sponsorship, affiliation or connection.” *Id.*

235. *See* Tony Lisanti, *The Rebound Continues*, LICENSE! MAG., Oct. 1, 2010, available at <http://www.licensemag.com/licensemag/Home/The-Rebound-Continues/ArticleStandard/>

Consumer search cost theory, which is grounded in consumer perception, is likely to lead to overprotection of trademark rights in the merchandising context, or at least to situations where we should seriously question whether trademark rights should be expanded.²³⁶

2. *Dilution.*—As noted previously, mark owners expend a tremendous amount of resources adopting and maintaining their marks. Marks are valued in the billions of dollars. It is not surprising that mark holders desire to protect extremely valuable marks even where there is no likelihood of confusion. Indeed, mark holders want to exploit their marks to the fullest extent possible, which includes garnering any revenue streams from the use of their marks, as well as ensuring that their marks continue to maintain their singularity as indications of source and their carefully crafted images. Thus, the antidilution cause of action, which does not require a demonstration of a likelihood of confusion, provides a powerful weapon for trademark owners. Courts and commentators have attempted to justify antidilution laws (under both a tarnishment and a blurring theory) with the consumer search cost theory.²³⁷ Judge Posner has provided an explanation of how antidilution law reduces consumer search costs:

[T]here is concern that consumer search costs will rise if a trademark becomes associated with a variety of unrelated products. Suppose an upscale restaurant calls itself “Tiffany.” There is little danger that the consuming public will think it’s dealing with a branch of the Tiffany jewelry store if it patronizes this restaurant. But when consumers next see the name “Tiffany” they may think about both the restaurant and the

Article/detail/690494.

236. Cf. Dogan & Lemley, *Merchandising Right*, *supra* note 67, at 463-65 (arguing generally against a broad merchandising right, for using disclaimers to avoid confusion, and against protection where the justification for liability is that it is property of the trademark owner). Notably, another area of expansion has occurred in cases involving sponsorship or association confusion. See generally Lemley & McKenna, *supra* note 99.

237. See, e.g., *Ty, Inc. v. Perryman*, 306 F.3d 509, 511 (7th Cir. 2002); LANDES & POSNER, *ECONOMIC STRUCTURE*, *supra* note 10, at 206-09; Sarah B. Chopnick, *Search Costs and Famous Foreign Marks: Should Congress Reduce the Search Costs of the Global Consumer and Protect Famous Foreign Marks?*, 5 SETON HALL CIRCUIT REV. 213, 215-16 (2008) (arguing that the blurring type of dilution is necessary to reduce consumer search costs.); Stacey L. Dogan & Mark A. Lemley, *What the Right of Publicity Can Learn from Trademark Law*, 58 STAN. L. REV. 1161, 1197 (2006) (“But like traditional trademark law, dilution properly understood is targeted at reducing consumer search costs.”); Richard A. Posner, *When Is Parody Fair Use?*, 21 J. LEGAL STUD. 67, 75 (1992) (“The idea is that the reputation built up by Coca-Cola may be impaired by the association of its trademarks with shoddy or sleazy activities, such as the traffic in illegal drugs. The idea can be given an economic form. A trademark seeks to economize on information costs by providing a compact, memorable, and unambiguous identifier of a product or service. The economy is less when, because the trademark has other associations, a person seeing it must think for a moment before recognizing it as the mark for the product or service.”).

jewelry store, and if so the efficacy of the name as an identifier of the store will be diminished. Consumers will have to think harder—incur as it were a higher imagination cost—to recognize the name as the name of the store.²³⁸

However, several commentators have criticized the use of the consumer search cost theory to support antidilution laws, particularly in the case of blurring.²³⁹ Indeed, blurring has come under frequent attack by commentators;²⁴⁰ some experts have tried to defend the blurring concept under the search costs rationale with cognitive theory by arguing that blurring leads to a delay in recall of a single source associated with a mark.²⁴¹ At least two commentators have argued that dilution should be justified under an anti-free riding rationale or perhaps a justification based on providing an incentive to adopt and maintain famous marks.²⁴² Implicitly, if not expressly, these authors recognize that search cost theory provides an inadequate basis for blurring or tarnishment theories of dilution, at least where there is unlikely to be any consumer confusion.²⁴³

The search cost theory appears to provide a weak basis for antidilution laws, which may help explain why the theory has not been a helpful limiting guide in determining when dilution should be actionable. Professor David Franklyn importantly notes:

There is scant empirical evidence that multiple uses of a famous mark dilute the selling power of the mark in connection with the first class of products to which it was attached. . . . Indeed, if dilution were a real risk, famous mark owners would rarely, if ever, license their marks for use in

238. *Perryman*, 306 F.3d at 511.

239. Robert G. Bone, *A Skeptical View of the Trademark Dilution Revision Act*, 11 INTELL. PROP. L. BULL. 187, 189-93 (2007); Robert G. Bone, *Schechter's Ideas in Historical Context and Dilution's Rocky Road*, 24 SANTA CLARA COMPUTER & HIGH TECH. L.J. 469, 473-74 (2008); see also Beebe, *Antidilution Law*, *supra* note 205, at 464 (“There are several fundamental problems with . . . [the search cost] theory of antidilution law, not the least of which is how it accounts for dilution by ‘tarnishment’ or why it should be provided only to famous marks.” (internal citation omitted)); Daniel Klerman, *Trademark Dilution, Search Costs, and Naked Licensing*, 74 FORDHAM L. REV. 1759, 1767 (2006) (“If blurring increases search costs only a little, the costs of the blurring cause of action may outweigh its benefits.”).

240. See generally Sara Stadler Nelson, *The Wages of Ubiquity in Trademark Law*, 88 IOWA L. REV. 731 (2003) (arguing that trademark owners who use their marks on diverse products—ubiquitous use—are not entitled to a remedy under dilution); Rebecca Tushnet, *Gone in Sixty Milliseconds: Trademark Law and Cognitive Science*, 86 TEX. L. REV. 507 (2008).

241. Jacob Jacoby, *The Psychological Foundations of Trademark Law: Secondary Meaning, Genericism, Fame, Confusion and Dilution*, 91 TRADEMARK REP. 1013, 1046-50 (2001); Jerre B. Swann, Sr., *Dilution Redefined for the Year 2002*, 92 TRADEMARK REP. 585, 606-23 (2002).

242. See Franklyn, *supra* note 82, at 117 (arguing for a dilution cause of action based on anti-free-riding rationale); Rose, *supra* note 82, at 663 (noting that antidilution rights are based on an earned property right theory).

243. See Franklyn, *supra* note 82, at 130.

collateral markets on a variety of different types of goods. Yet this type of licensing occurs frequently.²⁴⁴

Moreover, some arguably diluting uses may increase the source-identifying power of a mark, rather than dilute the mark. Thus, courts applying antidilution law are likely relying on anti-free riding concerns.

If consumer search cost theory did support antidilution law (and some believe it does), mark holders would continually push for greater rights to the detriment of the public interest, especially since they would be (at least in part) in control of the scope of their rights under a legal regime that focuses on consumer perception. However, this does not mean that mark holders should be entitled to control all uses of their marks, particularly those uses that impact the interests of the public, including the protection of other values. For example, a potentially tarnishing use should not be prohibited at the preliminary or permanent injunction stage if free speech values are implicated and outweigh other concerns. Given the reliance by some courts on a consumer search cost rationale for antidilution laws, it is not surprising that dilution law continues to expand despite efforts to rein it in.²⁴⁵

3. *New Theories of Confusion.*—Another area of expansion involves new theories of confusion: post-sale, reverse, initial interest, and sponsorship and association confusion. Under the consumer search cost theory, any likelihood of confusion at any conceivable stage in the purchase or use of a good or service could increase consumer search costs from the perspective of the consumer.

a. *Post-sale confusion.*—Post-sale confusion is confusion that occurs after the point of sale.²⁴⁶ For example, if the consumer encounters a mark in connection with a good that a person is wearing, such as a chevron-type symbol on the back of blue jeans, the consumer may be confused by a similarly shaped

244. *Id.* at 131 (internal citations omitted).

245. See *supra* notes 186-205 and accompanying text.

246. One of the first cases applying the post-sale confusion theory was *Mastercrafters Clock & Radio Co. v. Vacheron & Constantin-LeCoultre Watches, Inc.*, 221 F.2d 464, 466-67 (2d Cir. 1955) (finding likely confusion after the point of sale when visitors to owners of junior user's clock would think the clock was the senior user's more expensive Atmos clock). Many other courts have recognized post-sale confusion as a viable theory to support a finding of trademark infringement. See, e.g., *Karl Storz Endoscopy-Am., Inc. v. Surgical Techs., Inc.*, 285 F.3d 848, 854-55 (9th Cir. 2002) (holding that the district court erred in finding a lack of genuine issue of material fact in post-sale confusion); *Ferrari S.P.A. v. Roberts*, 944 F.2d 1235, 1244-45 (6th Cir. 1991) (finding likely confusion after point of sale based on sale of Ferrari replicas); *Keds Corp. v. Renee Int'l Trading Corp.*, 888 F.2d 215, 222 (1st Cir. 1989) (noting that prospective customers may be confused by a similar label on backs of shoes); *A.T. Cross Co. v. Jonathan Bradley Pens, Inc.*, 470 F.2d 689, 692 (2d Cir. 1972) (recognizing likely confusion amongst users of pens); *Adidas-Am., Inc. v. Payless Shoesource, Inc.*, 546 F. Supp. 2d 1029, 1058 (D. Or. 2008) (recognizing that initial interest and post-sale confusion are well established theories in the Ninth Circuit); *T. Anthony, Ltd. v. Louis Vuitton Malletier, No. 93 Civ. 6900*, 1993 WL 659682, at *4 (S.D.N.Y. Nov. 24, 1993) (recognizing likely confusion based on post-sale confusion theory).

symbol on another producer's blue jeans.²⁴⁷ The presence of additional labeling at the point of sale is irrelevant.²⁴⁸ Here, the consumer search cost theory would support a finding of confusion, likely under the rubric of a type of post-sale confusion. A consumer's search costs are higher after encountering post-sale confusion if the consumer attempts to find the same good that the consumer saw at an earlier time. A possibility also exists that consumers will associate the allegedly infringing mark with low quality goods.²⁴⁹ The question is whether trademark law should be concerned with this particular type of confusion or if it should only be concerned with confusion at the first point of sale. Consumer search cost theory would apparently make this type of confusion actionable, but how are the limits determined under the theory? And what about other interests—how are they to be taken into account?

In *General Motors Corp. v. Keystone Automotive Industries, Inc.*, the Sixth Circuit examined some of the possible harms that can arise from post-sale confusion involving knock-offs and the interests that should be balanced against those harms.²⁵⁰ The court stated that some of the recognized harms and interests involved in post-sale confusion included:

- (1) the viewing public, as well as subsequent purchasers, may be deceived if expertise is required to distinguish the original from the counterfeit; (2) the purchaser of an original may be harmed if the widespread existence of knockoffs decreases the original's value by making the previously scarce commonplace; (3) consumers desiring high quality products may be harmed if the original manufacturer decreases its investment in quality in order to compete more economically with

247. Lois Sportswear, U.S.A., Inc. v. Levi Strauss & Co., 799 F.2d 867, 872-73 (2d Cir. 1986) ("In the instant case, this post-sale confusion would involve consumers seeing appellant's jeans outside of the retail store, perhaps being worn by a passer-by. The confusion the Act seeks to prevent in this context is that a consumer seeing the familiar stitching pattern will associate the jeans with appellee and that association will influence his buying decisions."); Levi Strauss & Co. v. Blue Bell, Inc., 632 F.2d 817, 822 (9th Cir. 1980) ("Wrangler's use of its projecting label is likely to cause confusion among prospective purchasers who carry even an imperfect recollection of Strauss's mark and who observe Wrangler's projecting label after the point of sale. It is axiomatic in trademark law that 'side-by-side' comparison is not the test.'").

248. *Blue Bell, Inc.*, 632 F.2d at 822 ("[B]illboards and other point of sale materials are removed by the purchaser and have no confusion-obviating effect when the pants are worn.").

249. See *Polo Fashions, Inc. v. Craftex, Inc.*, 816 F.2d 145, 148 (4th Cir. 1987) (noting that in the post-sale context, customers may be confused by a mark and the plaintiff's reputation may be harmed if the mark is used in connection with apparently poor quality products); *Nabisco Brands, Inc. v. Conusa Corp.*, 722 F. Supp. 1287, 1291-92 (M.D.N.C. 1989) (finding post-sale confusion likely because of the risk of association with poor quality products); *Rolex Watch U.S.A., Inc. v. Canner*, 645 F. Supp. 484, 495 (S.D. Fla. 1986) (recognizing that after a product enters commerce, there may be confusion after the point of sale, including a "cheapening" of the products sold in connection with the mark).

250. *Gen. Motors Corp. v. Keystone Auto. Indus., Inc.*, 453 F.3d 351, 358 (6th Cir. 2006).

less expensive knockoffs; (4) the original manufacturer's reputation for quality may be damaged if individuals mistake an inferior counterfeit for the original; (5) the original manufacturer's reputation for rarity may be harmed by the influx of knockoffs onto the market; and (6) the original manufacturer may be harmed if sales decline due to the public's fear that what they are purchasing may not be the original. On the other hand, courts should be wary of overprotecting public domain ideas and works whose exploitation can lead to economic efficiency, greater competition, and lower costs for consumers.²⁵¹

The court ultimately decided that the alleged infringer's use of the plaintiff's marks could damage the plaintiff's reputation for producing quality goods.²⁵² However, while the court mentioned the public interests, it failed to engage in an analysis of whether those interests should dictate if post-sale confusion was found to exist or not.²⁵³ Notably, the court failed to consider any other public interest besides those closely associated with the consumer search cost theory. The consumer search cost theory based on consumer perception justifies providing a remedy for most, if not all, of the harms discussed by the court and does not inherently limit the scope of trademark liability in post-sale confusion cases.²⁵⁴

b. *Reverse confusion.*—Reverse confusion is a theory of trademark infringement that recognizes a likelihood of confusion where a junior user saturates the market with advertising, resulting in consumers associating the senior user's mark with the junior user's mark.²⁵⁵ In those circumstances, there is a likelihood of confusion, and the consumer search cost theory would support a finding of infringement. Consumers may be confused at the point of sale in determining the first user of the mark.²⁵⁶ This form of confusion is perhaps not as controversial as post-sale, initial interest, or sponsorship and association confusion. However, in a case based on post-sale, initial interest, or sponsorship and association confusion, reverse confusion could be troubling because of the problems with those forms of confusion and the consumer search cost theory.

c. *Initial interest confusion.*—Another expansion of confusion outside the

251. *Id.* (internal citations omitted).

252. *Id.* at 358-59.

253. *Id.*

254. McKenna, *Trademark Use*, *supra* note 14, at 821.

255. Big O Tire Dealers, Inc. v. Goodyear Tire & Rubber Co., 561 F.2d 1365, 1372 (10th Cir. 1977) (recognizing the reverse confusion theory); RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 20 cmt. f (1995) ("[I]n some cases the subsequent user's promotion of the mark may so overwhelm the use by the prior user that most purchasers come to associate the mark with the subsequent user. This can result in reverse confusion—purchasers are likely to believe that the goods sold by the prior user are actually those of the subsequent user. Reverse confusion is actionable . . .").

256. See *infra* notes 274-76. For an additional discussion of reverse confusion, see GRAEME B. DINWOODIE & MARK D. JANIS, TRADEMARKS AND UNFAIR COMPETITION 564-73 (2d ed. 2007).

point of sale is initial interest confusion.²⁵⁷ Initial interest confusion occurs when a particular party is confused not at the point of sale, but before the point of sale.²⁵⁸ The classic example involves a person driving on a freeway who sees a billboard sign stating that a McDonald's is at the next exit. When the person takes the exit, she discovers only a Burger King, without a McDonald's in sight. Having already exited (and desiring a hamburger), she goes ahead and eats at Burger King. Although the person is not confused that Burger King is McDonald's, she was initially confused, and this confusion resulted in her

257. See *supra* note 136 and accompanying text; see also *Brookfield Commc'n's v. W. Coast Entm't Corp.*, 174 F.3d 1036, 1063-65 (9th Cir. 1999) (listing cases discussing actionable initial interest confusion); *Mobil Oil Corp. v. Pegasus Petroleum Corp.*, 818 F.2d 254, 260 (2d Cir. 1987) ("[T]he district court's concerns focused upon the probability that potential purchasers would be misled into an initial interest in Pegasus Petroleum. Such initial confusion works a sufficient trademark injury."); *Adidas-Am., Inc. v. Payless Shoesource, Inc.*, 546 F. Supp. 2d 1029, 1058 (D. Or. 2008) (recognizing that initial interest and post-sale confusion are well-established theories in the Ninth Circuit). The doctrine of initial interest confusion was likely first raised in *Grotian, Helfferich, Schulz, Th. Steinweg Nachf. v. Steinway & Sons*, 365 F. Supp. 707 (S.D.N.Y. 1973). The *Grotian* court stated:

Misled into an initial interest, a potential Steinway buyer may satisfy himself that the less expensive Grotian-Steinweg is at least as good, if not better, than a Steinway. Deception and confusion thus work to appropriate defendant's good will. This confusion, or mistaken beliefs as to the companies' interrelationships, can destroy the value of the trademark which is intended to point to only one company. Thus, the mere fact that purchasers may be sophisticated or discriminating is not sufficient to preclude the likelihood of confusion. "Being skilled in their own art does not necessarily preclude their mistaking one trademark for another when the marks are as similar as those here in issue, and cover merchandise in the same general field." Even a discriminating purchaser might well assume that the marks in issue were trademarks on companion products of a single producer.

Id. at 717 (internal citations omitted).

258. See Dogan & Lemley, *Search Costs*, *supra* note 65, at 814-15 ("Initial interest confusion" as originally conceived did not reflect a new doctrine; rather, it was a simple recognition that competition-distorting confusion can occur at times other than the point of sale. Moreover, in the pre-Internet cases relying on the theory, the defendant had branded its product with a mark confusingly similar to the plaintiff's and had no comparative advertising or other pro-informational justification for its choice of mark. Given the net harm to consumers, these cases justified judicial relief."); Greg Lastowka, *Google's Law*, 73 BROOK. L. REV. 1327, 1369-70 (2008) ("The doctrine of initial interest confusion shifts the focus of confusion analysis to at a time prior to the time of purchase. Initial interest confusion can be found to exist even if confusion was not present at the time of purchase."); see generally Ross D. Petty, *Initial Interest Confusion Versus Consumer Sovereignty: A Consumer Protection Perspective on Trademark Infringement*, 98 TRADEMARK REP. 757 (2008) (examining initial interest confusion and disclosure remedies for trademark infringement in the context of FTC consumer protection policy); Jennifer E. Rothman, *Initial Interest Confusion: Standing at the Crossroads of Trademark Law*, 27 CARDOZO L. REV. 105 (2005).

diversion to the Burger King.²⁵⁹

Recently, the Ninth Circuit applied this theory in the Internet context in *Brookfield Communications, Inc. v. West Coast Entertainment Corp.*²⁶⁰ In that case, the court found initial interest confusion created by the infringer's use of a mark in metatags.²⁶¹ Here, the consumer search cost theory arguably supported a finding of initial interest confusion. Finding liability reduces consumer confusion that otherwise might delay a consumer's ability to obtain the goods or services in which she was initially interested.²⁶² There is also a counterargument under which the consumer search cost theory may not support finding infringement in this context, which posits that permitting the use of metatags would facilitate providing consumers with additional information about various choices they can make.²⁶³ However, attempting to draw the line between a situation involving actionable confusion with an unlawful diversion of a consumer versus one involving permissible confusion that may ultimately lead to a satisfied consumer with more choices is difficult.²⁶⁴ That question also requires an intensely fact-based inquiry. The nature of that fact-based inquiry makes it difficult to resolve the dispute early in litigation.²⁶⁵ The consumer search cost theory does not provide an adequate basis on which to make determinations of liability because it rests on consumer perception, and any confusion may increase consumer search costs.²⁶⁶ It also fails to provide adequate limits to trademark liability. Thus, the theory tends to be overinclusive and leads to findings of infringement that are not necessary to protect consumers and may actually harm them.

259. See *Brookfield Commc'ns*, 174 F.3d at 1064 (discussing a similar example).

260. *Id.* at 1065; see also Rothman, *supra* note 258, at 107-08.

261. *Brookfield Commc'ns*, 174 F.3d at 1065.

262. See Paul L. Bonevitz, Note, *Beyond Confusion: Reexamining Trademark Law's Goals in the World of Online Advertising*, 81 ST. JOHN'S L. REV. 899, 919 (2007) ("Proscribing non-confusing trademark uses that interfere with consumers reaching their search objectives is only justified on a search cost rationale, however, if the consumer's search objective is to access the trademark owner.").

263. See Michael Grynberg, *The Road Not Taken: Initial Interest Confusion, Search Costs, and the Challenge of the Internet*, 28 SEATTLE U. L. REV. 97, 99 (2004) ("[P]ermitting initial interest confusion may also harm consumers. The class of initially confused consumers includes those who are specifically seeking a particular brand to the *exclusion* of others. They must expend extra effort to determine which product is which, and to find their preferred choice. For these consumers, initial interest confusion impedes the trademark's function of reducing consumer search costs. This perspective suggests that a balancing is possible: Courts should police initial interest confusion only when it produces greater harms than benefits." (emphasis in original)).

264. See Eric Goldman, *Deregulating Relevancy in Internet Trademark Law*, 54 EMORY L.J. 507, 566 (2005).

265. Cf. Dogan & Lemley, *Search Costs*, *supra* note 65, at 805 ("The trademark use requirement serves a gatekeeper function, limiting the reach of trademark law without regard to factual inquiry into consumer confusion.").

266. McKenna, *Trademark Use*, *supra* note 14, at 821.

An example of a trademark owner asserting overbroad trademark rights with the initial interest confusion theory is *Gibson Guitar Corp. v. Paul Reed Smith Guitars, LP.*²⁶⁷ In that case, the court found that the initial interest confusion doctrine did not apply to plaintiff Gibson's trademark that essentially consisted of a guitar shape.²⁶⁸ The court rejected Gibson's initial interest confusion argument—apparently accepted by the district court—that consumers viewing the defendant's guitars from a distance would believe they were Gibson's guitars.²⁶⁹ The court reasoned, in part, that competition-related concerns outweighed the need to protect Gibson's trademark from potential initial interest confusion:

Many, if not most, consumer products will tend to appear like their competitors at a sufficient distance. Where product shapes themselves are trademarked, such a theory would prevent competitors from producing even *dissimilar* products which might appear, from the far end of an aisle in a warehouse store, somewhat similar to a trademarked shape.²⁷⁰

In that case, the Sixth Circuit properly recognized the potential problems with an expansive view of the initial interest confusion doctrine.

Some courts appear to base their finding of initial interest confusion on a misappropriation of goodwill argument, apparently relying on diversion of consumers.²⁷¹ Courts may be attracted to that rationale because the confusion is ultimately dispelled even though consumers may acquire additional information.²⁷² Notably, some courts have declined to adopt the initial interest

267. 423 F.3d 539 (6th Cir. 2005).

268. *Id.* at 553.

269. *Id.* at 550 n.15, 551-52.

270. *Id.* at 552 (emphasis in original) (internal citation omitted). The court also noted that recognizing initial interest confusion under similar facts would make it very difficult to dispose of cases through summary judgment. *Id.* at 550 n.15 (“If our belief that *nearly all* product-shape trademark-holders will be able to show an issue of fact as to whether a competing product creates initial-interest confusion is correct, application of the initial-interest-confusion doctrine in the product-shape context would make it substantially easier for product-shape trademark-holders to survive a defendant’s summary-judgment motion than for plaintiffs alleging any other type of trademark infringement.”) (emphasis in original)).

271. See *supra* note 136 and accompanying text.

272. In discussing the trademark “use” requirement for infringement in the context of the purchase of keywords for search engines, some courts have noted that the way some search results are displayed may enhance consumer choice. Cf. *Designer Skin, LLC v. S & L Vitamins, Inc.*, 560 F. Supp. 2d 811, 820 (D. Ariz. 2008) (“In this Court’s view, there is a meaningful distinction between (1) using a mark to attract potential customers to a website that only offers products of the mark holder’s competitors and (2) using a mark to attract potential customers to a website that offers the mark holder’s genuine products as well as the products of competitors. As discussed above, in the latter situation no ‘bait and switch’ occurs.”); see also Dogan & Lemley, *Search Costs*, *supra* note 65, at 815 (“[In] the online context, in which switching costs are minimal,

confusion theory.²⁷³

d. Sponsorship and association confusion.—The final expansion in confusion theories is sponsorship and association confusion.²⁷⁴ In sponsorship

confusion is frequently speculative, and many defendants have persuasive arguments that their uses bring benefits to consumers.” (internal citation omitted)). Moreover, searching on the Internet may not involve the same level of costs as a search in the brick and mortar world. *See Margreth Barrett, Domain Names, Trademarks and the First Amendment: Searching for Meaningful Boundaries*, 39 CONN. L. REV. 973, 1014 (2007).

273. *See, e.g.*, *Lamparello v. Falwell*, 420 F.3d 309, 316 (4th Cir. 2005) (“[W]e have never adopted the initial interest confusion theory; rather, we have followed a very different mode of analysis, requiring courts to determine whether a likelihood of confusion exists by ‘examin[ing] the allegedly infringing use *in the context in which it is seen by the ordinary consumer.*’” (emphasis in original) (citation omitted)). In *Lamparello*, the Fourth Circuit stated that the initial interest confusion doctrine is limited in other circuits to cases where “one business’s use of another’s mark for its own financial gain.” *Id.* at 317. At least one court has rejected the Fourth Circuit’s attempt to constrict the initial interest confusion theory in that way. *See SMJ Grp., Inc. v. 417 Lafayette Rest. LLC*, 439 F. Supp. 2d 281, 290 (S.D.N.Y. 2006) (“Despite defendants’ lack of profit motivation, or the lack of competition between the parties, an individual who is handed one of defendants’ pamphlets is, at least initially, confused about the source of the pamphlet.”).

274. Courts first recognized trademark infringement actions between direct competitors. *See Lemley & McKenna, supra* note 99, at 422. However, courts began to find trademark infringement (or unfair competition) between parties who may not directly compete, but compete in related fields—such that a consumer may expect that the trademark owner may expand into the other field. *See, e.g.*, *Yale Elec. Corp. v. Robertson*, 26 F.2d 972, 974 (2d Cir. 1928); *Aunt Jemima Mills Co. v. Rigney & Co.*, 247 F. 407, 409-10 (2d Cir. 1917). Prior to Congress’s 1962 amendment to the Lanham Act, the infringement provision instructed courts to consider whether a purported infringement “is likely to cause confusion, or to cause mistake, or to deceive” purchasers as to the source of origin of such goods or services.” Pub. L. No. 87-772, 76 Stat. 769. The 1963 amendment struck the language “purchasers as to the source of origin of such goods or services,” which some courts interpreted as removing any requirement that confusion is limited to purchasers or to any kind of confusion. *See, e.g.*, *Syntex Labs., Inc. v. Norwich Pharmacal Co.*, 437 F.2d 566, 568 (2d Cir. 1971). Moreover, the 1946 version of 15 U.S.C. § 1125(a) provided:

Any person who shall affix, apply, or annex, or use in connection with any goods or services, or any container or containers for goods, a false designation of origin, or any false description or representation, including words or other symbols tending falsely to describe or represent the same, and shall cause such goods or services to enter into commerce, and any person who shall with knowledge of the falsity of such designation of origin or description or representation cause or procure the same to be transported or used in commerce or deliver the same to any carrier to be transported or used, shall be liable to a civil action by any person doing business in the locality falsely indicated as that of origin or in the region in which said locality is situated, or by any person who believes that he is or is likely to be damaged by the use of any such false description or representation.

15 U.S.C. § 1125(a) (1946).

Congress’s 1988 amendments to the Lanham Act changed § 1125(a) to state:

and association confusion cases, courts are not focused on confusion as to the source of the goods and services. Rather, they are concerned that confusion might exist amongst consumers as to whether there is some relationship between the mark holder and the alleged infringer (i.e., that the mark owner has perhaps allowed the alleged infringer to use the mark).²⁷⁵ The consumer search cost theory supports finding an infringement in these circumstances. If a consumer mistakenly believes that the alleged infringer and mark holder are associated, there is some likelihood of confusion as to sponsorship or approval. This confusion may impact the consumer's perception of the mark owner or his decision to purchase from the mark owner in the future.²⁷⁶

Any person who, on or in connection with any goods or services, or any container for goods, uses in commerce any word, term, name, symbol, or device, or any combination thereof, or any false designation of origin, false or misleading description of fact, or false or misleading representation of fact, which . . . is likely to cause confusion, or to cause mistake, or to deceive as to the affiliation, connection, or association of such person with another person, or as to the origin, sponsorship, or approval of his or her goods, services, or commercial activities by another person . . . shall be liable in a civil action by any person who believes that he or she is or is likely to be damaged by such act.

Pub. L. No. 100-667, 102 Stat. 3935 (codified at 15 U.S.C. § 1125(a) (1988)). The amendments specifically changed § 1125(a) to include confusion as to sponsorship or approval of goods or services, thus adopting the cases which extended liability to sponsorship and approval confusion.

275. See, e.g., *Mut. of Omaha Ins. Co. v. Novak*, 836 F.2d 397, 403 (8th Cir. 1987) (upholding a preliminary injunction against the defendant's use of "Mutant of Omaha" based on the district court's finding of a likelihood of confusion as to Mutual of Omaha's sponsorship or affiliation with the defendant); see also Lemley & McKenna, *supra* note 99, at 416-22 (discussing examples of the expansion of actionable confusion in sponsorship and association cases); Malla Pollack, *Types of Trademark Infringement*, 160 INTELL. PROP. COUNSELOR 1 (2010) (discussing endorsement and association confusion cases).

276. This form of actionable confusion has recently been criticized by Professors Lemley and McKenna. See Lemley & McKenna, *supra* note 99, at 413-16 (positing that some courts have overreached with actionable sponsorship and affiliation confusion and that actionable confusion should be limited to confusion relevant to purchasing decisions). Professors Lemley and McKenna argue that

[t]he actionable confusion . . . [in some sponsorship and affiliation cases] was not confusion that would have led consumers to buy the wrong product, or even to wrongly think they were buying from the trademark owner. Rather, the theory in all of these cases was that consumers would think there was some relationship between the trademark owner and the defendant based on the defendant's use of the trademark. The problem with this formulation is that it fails to specify the types of relationships about which confusion is relevant or the harm that supposedly flows from confusion about those relationships. It is therefore impossible to establish meaningful limits on what sorts of confusion are actionable.

Id. at 421-22. For a discussion of the marketing literature concerning brand extension, see generally Lemley & McKenna, *supra* note 99; McKenna, *Theory of Harm*, *supra* note 58. Notably,

4. *Trademark Use.*—One debate confronting courts throughout the world concerns the role of trademark use in limiting what may constitute infringement or dilution.²⁷⁷ Scholars have intensely debated the question of whether use is and should be a limiting doctrine. One scholar has argued that trademark use is a historical requirement of trademark infringement; others argue that trademark use should be a requirement for infringement because keyword advertising increases the amount of information available to consumers, thereby decreasing search costs and providing a useful tool to limit trademark actions early.²⁷⁸ Other scholars vigorously argue that trademark use should not be used as a limiting doctrine because the concept of “use” in trademark law is anything but clear, and there could be consumer deception in cases involving keywords and Internet searching.²⁷⁹

The consumer search cost theory also operates to provide a justification for

the literature concerning brand extension generally appears not to support a theory of sponsorship or association confusion in unrelated markets. *See generally* Lemley & McKenna, *supra* note 99.

277. *See generally* Barrett, *Digital Technologies*, *supra* note 12; Margreth Barrett, *Finding Trademark Use: The Historical Foundation for Limiting Infringement Liability to the Uses “In the Manner of a Mark,”* 43 WAKE FOREST L. REV. 893 (2008) [hereinafter Barrett, *Finding Trademark Use*]; Margreth Barrett, *Internet Trademark Suits and the Demise of “Trademark Use,”* 39 U.C. DAVIS L. REV. 371 (2006); Stacey L. Dogan & Mark A. Lemley, *Grounding Trademark Law Through Trademark Use*, 92 IOWA L. REV. 1669 (2007); Dogan & Lemley, *Search Costs*, *supra* note 65, at 805 (“The trademark use requirement serves a gatekeeper function The rationale for the doctrine stems from the practical reality that it would be both unwise and impossible to permit trademark owners to control every use of their marks.”); Stacey L. Dogan & Mark A. Lemley, *The Trademark Use Requirement in Dilution Cases*, 24 SANTA CLARA COMPUTER & HIGH TECH. L.J. 541 (2008); Goldman, *supra* note 264. *But see* William G. Barber, *Dumping the “Designation of Source” Requirement from the TDRA: A Response to the Alleged “Trademark Use Requirement in Dilution Cases,”* 24 SANTA CLARA COMPUTER & HIGH TECH. L.J. 559 (2008) (arguing that the TDRA does not expressly include a use requirement and was not intended to include a use requirement); Dinwoodie & Janis, *Contextualism*, *supra* note 151, at 1641-42 (disagreeing with the opinion that the trademark use requirement would provide certainty); Graeme B. Dinwoodie & Mark D. Janis, *Lessons from the Trademark Use Debate*, 92 IOWA L. REV. 1703 (2007) [hereinafter Dinwoodie & Janis, *Trademark Use Debate*]; McKenna, *Trademark Use*, *supra* note 14, at 773 (“[C]ourts can determine whether a defendant has made trademark use of a plaintiff’s mark only by asking whether consumers are likely to view the defendant’s use as one that indicates the source of the defendant’s products or services. Because such an inquiry is, by its nature, highly context-sensitive, trademark use is not a concept capable of serving the limited function advocates hope.”).

278. *See* Barrett, *Digital Technologies*, *supra* note 12, at 8-10; Barrett, *Finding Trademark Use*, *supra* note 277, at 893-95; Dogan & Lemley, *Search Costs*, *supra* note 65, at 811 (“The trademark use doctrine strikes that balance in favor of permitting uses that are generally cost-reducing because they provide relevant, nonconfusing information to consumers, even if on occasion they may be cost-enhancing.” (internal citations omitted)).

279. *See* Dinwoodie & Janis, *Contextualism*, *supra* note 151, at 1641-42; Dinwoodie & Janis, *Trademark Use Debate*, *supra* note 277, at 1704-05.

a “trademark use” requirement. Some courts have relied upon the trademark use doctrine to dispose of Internet keyword suits, finding that purchasing keywords from search engines is not a trademark use and thus removing the opportunity to pursue infringement based solely on the purchase of trademarks as keywords. The Second Circuit addressed the role of trademark use as a prerequisite to an infringement action in the context of search engines selling keywords that correspond to trademarks.²⁸⁰ In that case, the court did not find trademark use as a barrier to an infringement claim based on the sale of keywords.²⁸¹ This interpretation may reduce consumer search costs where consumers are looking for a specific trademarked good or service, but it may take away the ability of consumers to have additional choices presented to them—ultimately increasing the costs of an ongoing search for anything other than a specific trademarked good or service. However, judging whether a consumer is still generally searching for a good or has made up her mind about the particular good she wants is very difficult.²⁸² Thus, the consumer search cost theory provides little aid in defining limits to trademark law, at least in the context of purchasing keywords from search engines.²⁸³

5. *The Dawn Donut Rule.*—Limiting doctrines such as the *Dawn Donut* rule (prohibiting relief against an alleged infringer in a distinct geographic market where no confusion is likely) have been eroded by courts in cases involving goods.²⁸⁴ The erosion of the *Dawn Donut* rule harms the ability of new entrants to establish themselves in markets when a competitor is operating in a remote market and has an Internet presence. Thus, allowing trademark holders to obtain injunctive relief prior to moving into a particular geographic market may allow

280. Rescuecom Corp. v. Google Inc., 562 F.3d 123 (2d Cir. 2009).

281. *Id.* at 131. The Ninth Circuit is following the Second Circuit’s decision. *See Network Automation, Inc. v. Advanced Sys. Concepts, Inc.*, No. 10-55840, 2011 WL 815806, at *4 (9th Cir. Mar. 8, 2011) (“We now agree with the Second Circuit that such use is a ‘use in commerce’ under the Lanham Act.”).

282. *See Goldman, supra* note 264, at 566 (“Searchers’ objectives cannot be inferred from the keywords they employ. . . . [I]t is improper to assume that using a trademarked keyword means that the searcher wanted to find the trademark owner. Because of objective opaqueness, any IIC doctrine built on diversion is inherently flawed. Finding searcher ‘diversion’ is not possible until one knows where searchers were heading in the first place.”).

283. Professors Dinwoodie and Janis argue against a trademark use requirement for infringement for several reasons, including the possibility of consumer deception in the use of trademarks in the search engine context. Dinwoodie & Janis, *Contextualism*, *supra* note 151, at 1626-28. Thus, there could be some search costs because of potential deception in the myriad of ways search results may be displayed presently and in the future. Professor Eric Goldman argues that there appears to be no basis to treat positive externalities provided by a trademark in the form of increased profits to third parties differently in the online and offline contexts. *See Eric Goldman, Brand Spillovers*, 22 HARV. J. L. & TECH. 381, 383-84 (2009). He “proposes to harmonize the legal treatment applicable to all intermediaries to encourage intermediaries to reduce consumer search costs, even if they profit from brand spillovers in doing so.” *Id.* at 384.

284. *See Cumbow, supra* note 64.

a competitor to protect its mark without adequately demonstrating success through sales under the mark in that market.²⁸⁵ Because of the mobility of consumers, the consumer search cost theory dismantles the *Dawn Donut* rule on the premise that consumers will travel from one part of the country to another and may be confused by similar marks on similar goods or services in geographically remote markets.²⁸⁶

6. *Overtaking the Public Domain.*—A potential area for additional trademark expansion is where trademark protection can extend to symbols or devices that are ordinarily protected by copyright and patent law and are now in the public domain.²⁸⁷ Trademark protection may subsist in a symbol or device as long as the mark is used in commerce. Thus, marks such as Mickey Mouse (and the character) may receive trademark protection beyond the copyright term, and the consumer search cost theory—in attempting to eliminate consumer confusion—would support that trademark protection.²⁸⁸ Here again, the search cost theory alone would fail to provide a meaningful limit to trademark rights, and other values and interests should be considered.

III. PROPOSALS FOR REFORM

The first proposal for reform considers both preliminary and permanent injunctive relief in trademark matters. The second proposal argues that policing efforts of mark holders should not be allowed as evidence to demonstrate whether a mark should receive protection or as evidence of the scope of that protection.

A. *The Public Interest, Values, and Injunctive Relief*

This section will discuss the interests of the various parties in the context of injunctive relief. Preliminary injunctions are discussed first, and permanent injunctions are reviewed second.²⁸⁹

285. See Carter, *supra* note 61, at 796 (arguing that the *Dawn Donut* rule limiting injunctions should be preserved to limit expanding trademark rights). But see David S. Welkowitz, *The Problem of Concurrent Use of Trademarks: An Old/New Proposal*, 28 U. RICH. L. REV. 315, 342-44, 384 (1994) (arguing that the *Dawn Donut* rule should be overturned to provide full effect to the Lanham Act's constructive notice provision).

286. For additional discussions of the *Dawn Donut* rule, see generally Thomas L. Casagrande, “*The Dawn Donut Rule*”: Still Standing (Article III, That Is) Even with the Rise of the Internet, 90 TRADEMARK REP. 723 (2000) and Jessica Amber Drew, *Death of Dawn Donut: The Demise of Concurrent Trademarks*, 2007 U. ILL. J.L. TECH. & POL’Y 145.

287. See Moffat, *supra* note 133, at 1474.

288. Courts in several cases have effectively policed the boundary between trademarks and copyright and patent law. See, e.g., *Dastar Corp. v. Twentieth Century Fox Film Corp.*, 539 U.S. 23 (2003); *TraffFix Devices, Inc. v. Mktg. Displays, Inc.*, 532 U.S. 23 (2001); *Fleischer Studios, Inc. v. A.V.E.L.A., Inc.*, No. 09-56317, 2011 WL 631449 (9th Cir. Feb. 23, 2011).

289. Section 1116(a) of 15 U.S.C., in pertinent part, provides:

The several courts vested with jurisdiction of civil actions arising under this chapter

1. *Preliminary Injunctions.*²⁹⁰—The first proposal for reform is to reinvigorate the public interest prong of the injunctive relief analysis by expressly requiring district courts to balance the various interests of the public—including users, consumers, and competitors—with the interests of the trademark holder in analyzing whether a preliminary injunction should issue. This requirement will make district court judges specifically take into account other interests besides consumer confusion and the protection of the goodwill of the trademark holder. This determination would also occur early enough to deny relief to trademark owners in certain circumstances—for example, allowing expression to continue during litigation. As more reported cases deny preliminary and permanent injunctions, trademark owners will have an incentive to settle early and not raise overreaching claims in the first place, eliminating some of the chilling effects provided by current trademark law and practice.²⁹¹ Although this proposal may not completely solve the inherent problems of trademark law, it moves us towards recognizing values and interests in trademark law and reconciling the interests at stake.

The Lanham Act specifically provides district courts discretion to issue injunctions where equity allows it.²⁹² Appellate courts have established various tests for determining whether a preliminary injunction should issue. Generally,

shall have power to grant injunctions, according to the principles of equity and upon such terms as the court may deem reasonable, to prevent the violation of any right of the registrant of a mark registered in the Patent and Trademark Office or to prevent a violation under subsection (a), (c), or (d) of section 1125 of this title.

15 U.S.C. § 1116(a) (2006).

290. The importance of preliminary injunctions in trademark cases cannot be understated. Professor McCarthy recently noted that

[a] preliminary injunction is one of the most powerful weapons in the arsenal of a trademark owner. Getting a preliminary injunction means that the trademark owner can force the alleged infringer to immediately stop all use of the challenged mark and undergo an expensive change to a significantly different mark. That change will last for the months or years that will ensue until all the issues can be hashed out in a full-fledged trial on the merits. In some situations, getting a preliminary injunction means that the trademark owner will immediately receive just about all the relief it would be entitled to even after a win on the merits at trial.

J. Thomas McCarthy, *Are Preliminary Injunctions Against Trademark Infringement Getting Harder to Achieve?*, 14 INTELL. PROP. L. BULL. 1, 1 (2009).

291. See Ramsey, *supra* note 38, at 384-85 (“Protected expression is frequently suppressed or chilled by trademark law because the law’s current built-in First Amendment safeguards, such as the descriptive fair use doctrine, are limited and involve fact-specific determinations that often can only be resolved after discovery at summary judgment or trial—a cost many defendants cannot afford.”); see also McGeveran, *Free Speech*, *supra* note 26, at 1206-07 (“Considerable anecdotal evidence suggests that the real action occurs outside the courthouse: markholders send cease-and-desist letters and threaten legal action against those using trademarks to facilitate speech, and the recipients frequently capitulate.”).

292. 15 U.S.C. § 1116.

the test will include four factors: (1) the likelihood of success on the merits; (2) the presence or absence of irreparable harm; (3) the balance of the hardships between plaintiff and defendant; and (4) the public interest. In trademark cases, courts have often presumed irreparable harm in cases where a likelihood of success on the merits has been demonstrated.²⁹³ The justifications for presuming irreparable harm are that it is particularly difficult to ascertain damages in trademark cases and that the trademark owner has lost control over its goodwill.²⁹⁴ Moreover, some courts are not even considering the public interest in determining whether a preliminary injunction should issue in trademark cases.²⁹⁵ In a 2009 decision, *Zino Davidoff SA v. CVS Corp.*, the Second Circuit

293. *Genesee Brewing Co. v. Stroh Brewing Co.*, 124 F.3d 137, 142 (2d Cir. 1997) (“In the context of trademark and unfair competition injunctions, the requirement of irreparable harm carries no independent weight, as we have held that a showing of likelihood of confusion (a requirement of both trademark infringement and unfair competition claims) establishes irreparable harm.”). However, the Eleventh and Fifth Circuits have questioned whether the presumption of irreparable harm applies post-*eBay* and may require proof of irreparable harm. *See Paulsson Geophysical Servs., Inc. v. Sigmar*, 529 F.3d 303, 312-313 (5th Cir. 2008) (discussing *eBay*); *N. Am. Med. Corp. v. Axiom Worldwide, Inc.*, 522 F.3d 1211, 1227-28 (11th Cir. 2008) (discussing *eBay*). Moreover, additional cracks in the presumption supporting irreparable harm are becoming visible after the U.S. Supreme Court’s decision in *Winter v. NRDC*. *See Mortg. Elec. Registration Sys., Inc. v. Brosnan*, No. C 09-3600 SBA, 2009 WL 3647125, at *8 (N.D. Cal. Sept. 4, 2009) (noting that the U.S. Supreme Court in *Winter* has effectively eliminated the presumption of irreparable harm based on a showing of probable success on the merits); *Cytosport, Inc. v. Vital Pharm., Inc.*, 617 F. Supp. 2d 1051, 1081 (E.D. Cal. 2009) (same), *aff’d*, 348 F. App’x 288 (9th Cir. 2009); *Volkswagen AG v. Verdier Microbus & Camper, Inc.*, No. C 09-00231 JSW, 2009 WL 928130, at *8 (N.D. Cal. Apr. 3, 2009) (same). Those courts are still finding irreparable harm based upon a showing of goodwill, the difficulty of determining damages, and the lack of control over the reputation of the mark by the mark owner. *Brosnan*, 2009 WL 3647125, at *8; *Cytosport*, 617 F. Supp. 2d at 1081; *Volkswagen AG*, 2009 WL 928130, at *8. At least one court has applied the presumption despite the *Winter* case. *See MoroccanOil, Inc. v. Moroccan Gold, Inc.*, 590 F. Supp. 2d 1271 (C.D. Cal. 2008). One author has argued that given the uncertainty in cases concerning the status of the presumption of irreparable harm post-*Winter*, mark owners should be prepared to demonstrate irreparable harm. *See Thomas M. Williams, Winter v. NRDC: A Stricter Standard of Irreparable Harm in Trademark Cases?*, 91 J. PAT. & TRADEMARK OFF. SOC’Y 571, 579 (2009).

294. *Helene Curtis Indus., Inc. v. Church & Dwight Co.*, 560 F.2d 1325, 1332-33 (7th Cir. 1977) (“Where there is . . . such high probability of confusion, injury irreparable in the sense that it may not be fully compensable in damages almost inevitably follows. While an injured plaintiff would be entitled to recover the profits on the infringing items, this is often difficult to determine; moreover, a defendant may have failed to earn profits because of the poor quality of its product or its own inefficiency. Indeed, confusion may cause purchasers to refrain from buying either product and to turn to those of either competitors.” (quoting *Omega Importing Corp. v. Petri-Kine Camera Co.*, 451 F.2d 1190, 1195 (2d Cir. 1971)).

295. *See Zino Davidoff SA v. CVS Corp.*, 571 F.3d 238, 242 (2d Cir. 2009); *Arrow United Indus., Inc. v. Hugh Richards, Inc.*, 678 F.2d 410, 413-14 (2d Cir. 1982). In *GoTo.com, Inc. v. Walt Disney Co.*, 202 F.3d 1199 (9th Cir. 2000), the Ninth Circuit did not apply the public interest

analyzed whether to grant a preliminary injunction in a trademark infringement matter without explicitly noting that the public interest was a factor to consider.²⁹⁶ However, even if a court does not expressly consider the public interest as a factor, it may do so implicitly. For example, in *Helene Curtis Industries, Inc. v. Church & Dwight Co.*, the court noted the “probable right of both plaintiff and the public under the law to relief against continuation of . . . [defendant’s] misleading, confusing and deceptive acts.”²⁹⁷ Thus, the court collapsed the rights of the public with those of the trademark holder, but it failed to analyze the public interest as a separate and distinct factor and discuss that interest in any depth.²⁹⁸

After the Supreme Court’s decision in *Winter v. Natural Resources Defense Council, Inc.*,²⁹⁹ courts should not engage in a similar analysis. *Winter* appears to require consideration of the traditional four factors for a preliminary injunction to issue in every case and demonstrate generally that a court’s presumption of irreparable harm may be disfavored.³⁰⁰ Indeed, in *Winter*, the Supreme Court

factor. A district court in the Ninth Circuit recently applied the public interest prong of the test, but only in the context of discussing documents that were inadvertently sent to another party—there was no discussion of the independent importance of the public interest in the trademark context. *See Brosnan*, 2009 WL 3647125, at *8. Professor McCarthy notes that at least before *Winter*, the Second and Ninth Circuits apparently do not explicitly consider the public interest, although other circuits appear to do so. *See* 5 MCCARTHY, *supra* note 173, at § 30:32. In *New York City Triathlon, LLC v. NYC Triathlon Club, Inc.*, 704 F. Supp. 2d 305 (S.D.N.Y. 2010), the court noted that *Winter* requires a consideration of the public interest, but the Second Circuit does not under *Zino Davidoff*. *Id.* at 313. However, the court did not choose between either standard; it avoided the question and held that both standards were met. *Id.* The court did not expressly apply the public interest factor required by *Winter*. *Id.*

296. *Zino Davidoff SA*, 571 F.3d at 242 (“In cases involving claims of trademark infringement and dilution, as in other types of cases, a party seeking a preliminary injunction must demonstrate (1) the likelihood of irreparable injury in the absence of such a injunction, and (2) either (a) likelihood of success on the merits or (b) sufficiently serious questions going to the merits to make them a fair ground for litigation plus a balance of hardships tipping decidedly toward the party requesting the preliminary relief.” (quoting *Fed. Express Corp. v. Fed. Espresso, Inc.*, 201 F.3d 168, 173 (2d Cir. 2000)). Other recent trademark cases in the Second Circuit have failed to raise the public interest prong. *See, e.g., Cold Stone Creamery, Inc. v. Gorman*, 361 F. App’x 282, 287 (2d Cir. 2010); *N.Y. State Elec. & Gas Corp. v. U.S. Gas & Elec., Inc.*, 697 F. Supp. 2d 415, 425 (W.D.N.Y. 2010).

297. *Helene Curtis Indus., Inc.*, 560 F.2d at 1330.

298. *See id.*

299. 129 S. Ct. 365, 375 (2008) (rejecting the Ninth Circuit’s “possibility of irreparable harm” standard and stating that there must be a likelihood of irreparable harm).

300. In a recent article, some commentators assessed the impact of *eBay* on the presumption of irreparable harm in trademark infringement cases:

[We conclude] that *eBay* should not be used to eviscerate the normal presumption of irreparable harm that attaches upon a showing of liability in trademark cases. Although trademark, patent and copyright laws are all forms of rights in intangible property, the

specifically cautioned courts to consider the public interest carefully before issuing a preliminary injunction.³⁰¹ The plaintiff in *Winter* sought to preliminarily enjoin the Navy from conducting sonar training that purportedly harmed marine wildlife.³⁰² The Ninth Circuit affirmed the district court's issuance of a preliminary injunction and held that once a "strong" likelihood of success is proved, a plaintiff only needs to demonstrate a possibility of irreparable harm—not likely irreparable harm—to obtain preliminary injunctive relief.³⁰³ The Supreme Court reversed the Ninth Circuit, stating that the "possibility" standard for proving irreparable harm was too lenient and that even if the plaintiffs were successful in showing irreparable harm, any irreparable harm was outweighed by the public interest in national security.³⁰⁴ The Supreme Court stated:

A preliminary injunction is an extraordinary remedy never awarded as of right. In each case, courts must balance the competing claims of injury and must consider the effect on each party of the granting or

rationales underlying trademark protection are sufficiently distinct from those motivating patent and copyright protections that principles applicable in one type of intellectual property law are not necessarily appropriate in another. Patent and copyright protections are designed to be a trade-off between incentivizing innovation and creativity on one hand, and public access to the fruits of such labor on the other. Trademark law does not entail such a direct balance between individual gain and the public interest; rather, it is a form of consumer and competitor protection that is designed to efficiently indicate the source of a product or service and to avoid confusion. Because maximization of the public interest in trademark protection rests upon a very different protection rationale than that underlying patent and copyright laws, courts should not hastily import the Supreme Court precedent addressing a markedly different set of public interest concerns into the trademark realm.

David H. Bernstein & Andrew Gilden, *No Trolls Barred: Trademark Injunctions After eBay*, 99 TRADEMARK REP. 1037, 1038-39 (2009). Professor McCarthy has similarly opined that the presumption of irreparable harm for trademark cases should not be upset by *eBay*. See McCarthy, *supra* note 290, at 3-4. For a contrary viewpoint, see Sandra L. Rierson, *IP Remedies After eBay: Assessing the Impact on Trademark Law*, 2 AKRON INTELL. PROP. J. 163, 166 (2008). For additional commentary, see Jonathan Hudis et al., *Why Trademark and Copyright Counsel Should Heed the Patent Precedent of the Supreme Court*, 2 LANDSLIDE 14 (2009).

301. *Winter*, 129 S. Ct. at 376. The Court opined, "In exercising their sound discretion, courts of equity should pay particular regard for the public consequences in employing the extraordinary remedy of injunction." *Id.* at 376-77 (quoting *Weinberger v. Romero-Barcelo*, 456 U.S. 305, 312 (1982)). At least one court has held that in the context of trademark cases, the presumption of irreparable harm may not be available for trademark cases post-*Winter*. See *Mortg. Elec. Registration Sys., Inc. v. Brosnan*, No. C 09-3600 SBA, 2009 WL 3647125, at *8 (N.D. Cal. Sept. 4, 2009).

302. *Winter*, 129 S. Ct. at 372-74.

303. *Id.* at 375.

304. *Id.* at 375-77.

withholding of the requested relief. In exercising their sound discretion, courts of equity should pay particular regard for the public consequences in employing the extraordinary remedy of injunction.³⁰⁵

Thus, the Supreme Court requires a consideration of the public interest by the district court before granting a preliminary injunction.

However, even courts analyzing the public interest have frequently defaulted to granting injunctive relief because of the possibility of consumer deception and fail to analyze other potential harms to the public interest.³⁰⁶ Courts may continue to engage in that practice even after *Winter*. Indeed, more than one court, while citing *Winter* and reciting the four factors for preliminary relief, failed to apply the public interest factor.³⁰⁷ Courts should follow *Winter* and consider the public interest in granting preliminary injunctions.³⁰⁸ Courts should also expressly consider each potential public interest in trademark cases.

In considering whether a preliminary injunction should issue, the public interest factors, including those favoring trademark owners and the public generally (including users, consumers, and competitors), should be expressly weighed by the court where applicable. The court should not rely on the “built-in” public interest protections in trademark law such as the protection of free expression. For example, free expression can be protected by doctrines such as the likelihood of confusion test, fair use defenses, or a First Amendment defense, but those doctrines may not adequately protect the interest in free expression.³⁰⁹

305. *Id.* at 376-77 (internal citations and quotations omitted).

306. GSC Logistics, Inc. v. Star Galaxy Logistics, Inc., No. C 09-5886 SBA, 2010 WL 690200, at *4 (N.D. Cal. Feb. 24, 2010) (“The harm to Plaintiff and the public resulting from Defendants’ continued use of the Plaintiff’s name is evident from the fact Defendant’s services have been used by customers believing that they, in fact, were dealing with Plaintiff.”); Nike, Inc. v. Nikepal Int’l, Inc., No. 2:05-cv-1468-GEB-JFM, 2007 WL 2782030, at *8 (E.D. Cal. Sept. 18, 2007) (discussing the public interest factor in granting permanent injunctions and stating that the public interest is served by allowing the mark to continue to serve a source identification function).

307. Edom Labs., Inc. v. Special Tea Plus, Inc., No. CV-09-5185, 2010 WL 596342, at *1-2 (E.D.N.Y. Feb. 16, 2010). In *Fendi Adele S.R.L. v. Ashley Reed Trading, Inc.*, No. 06 Civ. 243, 2010 WL 571804 (S.D.N.Y. Feb. 16, 2010), the court did not discuss the public interest factor in deciding to grant a permanent injunction. *Id.* at *5. In *New York City Triathlon, LLC v. NYC Triathlon Club, Inc.*, 704 F. Supp. 2d 305 (S.D.N.Y. 2010), the court noted the public interest factor under *Winter* but apparently failed to explicitly apply the factor.

308. In *Network Automation Inc. v. Advanced Systems Concepts, Inc.*, No. 10-55840, 2011 WL 815806, at *3 (9th Cir. Mar. 8, 2011), the Ninth Circuit cited the *Winter* four-part test for issuing a preliminary injunction in a trademark case. *Id.*

309. See Rebecca Tushnet, *Trademark Law as Commercial Speech Regulation*, 58 S.C. L. REV. 737, 743-44 (2007) (noting that many doctrines in trademark law except for nominative fair use “exist to balance consumer protection with free competition rather than to implement First Amendment norms”); see also *Facenda v. N.F.L. Films, Inc.*, 542 F.3d 1007, 1015-18 (3d Cir. 2008) (considering First Amendment defense); *Lamparello v. Falwell*, 420 F.3d 309, 313-15 (4th Cir. 2005) (discussing defenses to dilution actions and likelihood of confusion test that protect free

The primary concern, particularly with free expression, is that this expression will be chilled prior to the litigation. The prospect of a denial of a preliminary injunction may dissuade some trademark owners from moving beyond a cease and desist letter. Furthermore, this prospect may dissuade mark owners from sending such letters in the first place, particularly after there are reported decisions denying injunctive relief.

The public's interests, as discussed previously, can be unpacked to include consumer interests, user interests, and competitor interests. The first interest is the consumer's interest in preventing confusion and deception in the marketplace, thus enabling trademarks to foster the market economy.³¹⁰ The second interest is the user's interest in free expression and in the right of others to listen to that free expression. The third interest is the consumer's interest in access to comparative advertising information and the competitor's right to use a mark to fairly describe its goods and services and in comparative advertising. The fourth interest is the user's interest in protecting cultural heritage. Moreover, the court can also consider the interests of the mark owner, such as the investment in goodwill in the mark, business development, and the mark holder's role in bringing actions to dispel consumer confusion in the marketplace. Courts can identify other interests as well, but they should be required at a minimum to consider these interests (when applicable in the overall balance) with the other factors considered in granting a preliminary injunction.

Courts should be particularly cautious with the strong likelihood that free speech may be chilled at the preliminary injunction stage and lean toward not granting preliminary injunctions at that point.³¹¹ In balancing the factors in a case where free expression is implicated, the court should not issue a preliminary injunction unless there is an intent to deceive consumers,³¹² a demonstration of irreparable harm, and a clear likelihood of success on the merits, including perhaps strong evidence of actual confusion or dilution by survey.³¹³ This

speech interests).

310. This also includes the interest in ensuring that new competitors can enter the market.

311. *See Winter v. Natural Res. Def. Council, Inc.*, 129 S. Ct. 365, 376 (2008) (“A preliminary injunction is an extraordinary remedy never awarded as of right.” (citing *Munaf v. Geren*, 553 U.S. 674, 690 (2008))).

312. *Cf. Maker’s Mark Distillery, Inc. v. Diageo N. Am., Inc.*, 703 F. Supp. 2d 671, 701 (W.D. Ky. 2010) (noting that injunctive relief tailored to the infringement is favored especially where there is no evidence of defendant’s bad intent.).

313. In *SMJ Group, Inc. v. 417 Lafayette Restaurant LLC*, 439 F. Supp. 2d 281 (S.D.N.Y. 2006), the court found that the plaintiff would likely succeed on the merits by proving a likelihood of initial interest confusion, but the court ultimately declined to issue a preliminary injunction (even with a presumption of irreparable harm) because there was no irreparable harm. *Id.* at 293-95. The court determined, in part, that: there was no lingering confusion after the consumer was initially confused; there was no diversion of consumers from use of the mark; and the defendant was not engaged in a for-profit enterprise, so there was no need to engage in a complicated disgorgement analysis. *Id.* The court also noted that an injunction might interfere with First Amendment expression interests, but it did not expressly analyze those interests under the public interest prong.

restriction is particularly important where the primary theory underlying the trademark cause of action is based on a persuasive—as opposed to a source-identifying—function.³¹⁴ Moreover, in cases involving dilution, initial interest confusion, and merchandising, the interest in free speech should outweigh concerns about lowering consumer search costs depending on the facts of the case.³¹⁵ Courts have engaged in similar analyses in trademark cases, such as the balancing of interests the Second Circuit exercised in *Rogers v. Grimaldi*.³¹⁶ In *Rogers*, the plaintiff claimed that the defendant's film *Ginger and Fred* created the false impression that Rogers was associated with the film.³¹⁷ The Second Circuit stated that

the [Lanham] Act should be construed to apply to artistic works only where the public interest in avoiding consumer confusion outweighs the

Id. Notably, the court separated the defendant's use of the plaintiff's trademark on the sheet of paper visible to the consumer from the expressive material inside the pamphlet in refusing to grant an injunction. *Id.* Thus, even if the court had issued an injunction, the defendant could have continued to pass out leaflets without the plaintiff's mark criticizing the plaintiff's treatment of its workers because the injunction would likely be narrowly tailored to prohibit the defendant's use of the plaintiff's mark. In *Network Automation, Inc. v. Advanced Systems Concepts, Inc.*, No. 10-55840, 2011 WL 815806 (9th Cir. Mar. 8, 2011), the Ninth Circuit noted that the actual confusion factor is not as important at the preliminary injunction stage. *Id.* at *9.

314. See Heymann, *supra* note 20, at 656-58 (advocating a view of the consumer grounded in autonomy theory that recognizes that “the consumer be left free to make whatever associations she wants with the marks she encounters, even if those associations are not the ones the markholder would prefer, or not the ones that would be optimal from the perspective of the individual’s intellectual or personal development,” particularly as against theories of trademark law “that depend in part on the persuasive value of the mark rather than its source-identifying aspects” and that “recognition of consumers’ autonomy interests suggests that the farther the doctrine moves away from instances in which consumer confusion as to source is the harm to be prevented—the paradigmatic trademark case—the less vigorous a role trademark law should play”); *see generally* Ralph S. Brown, Jr., *Advertising and the Public Interest: Legal Protection of Trade Symbols*, 57 YALE L.J. 1165 (1948).

315. See Heymann, *supra* note 20, at 659 (“In each of these areas [dilution, initial interest confusion, and merchandising], trademark law intervenes to limit the consumer’s decision making in response to the persuasive value of the mark, channeling consumers’ mental associations with the mark and thus impinging on autonomy that is necessary to discourse and personal development.”).

316. 875 F.2d 994, 998-1002 (2d Cir. 1989). Several other cases have considered or applied the *Rogers* test. *See, e.g.*, *Facenda v. N.F.L. Films, Inc.*, 542 F.3d 1007, 1015-18 (3d Cir. 2008); *Parks v. LaFace Records*, 329 F.3d 437, 449 (6th Cir. 2003); *Mattel, Inc. v. MCA Records, Inc.*, 296 F.3d 894, 901-02 (9th Cir. 2002); *Cliffs Notes, Inc. v. Bantam Doubleday Dell Publ’g Grp., Inc.*, 886 F.2d 490, 493-97 (2d Cir. 1989), *superseded by statute*, 15 U.S.C. §§ 1114(a), 1125(a)(1)(A) (2006); *Volkswagen AG v. Dorling Kindersley Publ’g, Inc.*, 614 F. Supp. 2d 793, 809-11 (E.D. Mich. 2009).

317. *Rogers*, 875 F.2d at 1000-01.

public interest in free expression. In the context of allegedly misleading titles using a celebrity's name, that balance will normally not support application of the Act unless the title has no artistic relevance to the underlying work whatsoever, or, if it has some artistic relevance, unless the title explicitly misleads as to the source or the content of the work.³¹⁸

Courts should similarly balance the trademark owner's interest as well as the public's relative interests under the public interest prong in determining whether to issue preliminary relief. Courts could also consider not only the particular title of the work and whether it was used in a work of artistic expression, but also the type of confusion alleged and its cost to the various public interests. Moreover, courts could consider the interest of the public in not chilling free expression at this stage of the litigation. Courts could deny issuing a preliminary injunction even if under a *Rogers* analysis, the use of a title might be found to cause a likelihood of confusion that outweighs the public interest in free speech. An example may include the use of a title in connection with political speech.³¹⁹

If other interests are involved, courts must expressly consider those interests. To the extent that the user exercises its cultural rights, this interest should be taken into account and weighed against the mark holder's interest in protecting its goodwill.³²⁰ If a court finds sufficient proof of irreparable injury along with a likelihood of success on the merits and decides to issue a preliminary injunction, the scope of that injunction should be as narrow as possible to protect the public's interest.³²¹ For example, courts may authorize the use of disclaimers

318. *Id.* at 999 (citation omitted).

319. See *Mastercard Int'l Inc. v. Nader* 2000 Primary Comm., Inc., No. 00 Civ. 6068, 2004 WL 434404, at *4 (S.D.N.Y. Mar. 8, 2004) (finding "no genuine issue of material fact with regard to any likelihood of confusion between Mastercard's Priceless Advertisements and Ralph Nader's Political Ad").

320. See Carpenter, *supra* note 148, at 929 ("[I]ndigenous peoples' rights . . . [must be protected] to enjoy certain cultural rights respecting cultural participation, identity, and self-determination. One of the most apparent examples of such a conflict involves the appropriation of aspects of indigenous cultural identity to signify the goods and services of non-indigenous, commercial interests.").

321. Notably, Professor McKenna has argued that lately, courts have issued broad injunctive relief as opposed to limited injunctive relief and that a proper exercise of equitable discretion by courts (such as by using disclaimers), could mitigate some of the harms caused by overreaching trademark law in cases involving mixed consumer understanding, sponsorship or affiliation confusion, merchandising cases, and protecting speech values. See Mark P. McKenna, *Back to the Future: Rediscovering Equitable Discretion in Trademark Cases*, 14 LEWIS & CLARK L. REV. 537, 544-52 (2010). According to Professor McKenna, an effective use of equitable discretion may allow potential alleged infringers to better predict whether their use will be infringing or not. *Id.* at 552-53. In *Toyota Motor Sales, U.S.A., Inc. v. Tabari*, 610 F.3d 1171 (9th Cir. 2010), the Ninth Circuit noted that "[a] trademark injunction, particularly one involving nominative fair use, can raise serious First Amendment concerns because it can interfere with truthful communication between buyers and sellers in the marketplace." *Id.* at 1176.

to mitigate any possible confusion or dilution.³²²

Although a bond may be sufficient to protect the interests of the party that is enjoined, it is generally insufficient to protect the public's interest and the harm from the chilling of speech.³²³ The bond covers the costs and pecuniary injury of a wrongfully enjoined party, which likely will not include a consideration of the costs to the public.³²⁴ This proposal also addresses concerns raised by Professor Michael Grynberg—notably, that formalism may hinder the development of defenses adequately protecting important values impacted by trademark law, such as free speech.³²⁵

2. *Permanent Injunctions.*—The recent Supreme Court patent case *eBay Inc. v. MercExchange, L.L.C.* supports the proposal for courts to consider the public interest and balance each relevant interest against the other factors used to determine if injunctive relief should issue in trademark cases.³²⁶ In *eBay*, the Court specifically required balancing four factors in determining whether permanent injunctive relief should issue when a federal statute provides the court with equitable powers to issue an injunction in its discretion.³²⁷ Although that decision specifically concerned patent law (and the Court mentioned copyright law in its opinion), the same general principles in equity should also apply to trademark law. The *eBay* Court provided four factors that should be considered before granting permanent injunctions:

A plaintiff must demonstrate: (1) that it has suffered an irreparable injury; (2) that remedies available at law, such as monetary damages, are inadequate to compensate for that injury; (3) that, considering the balance of hardships between the plaintiff and defendant, a remedy in equity is warranted; and (4) that the public interest would not be

322. Cf. 5 MCCARTHY, *supra* note 173, at § 30:3 (“In litigation between persons of the same personal name using the name as a trademark, it is common practice to balance the rights by a limited injunction which, while not preventing the use of a personal name altogether, requires some change in format, the addition of prefixes or suffixes, or disclaimers.” (citations omitted)).

323. See FED. R. CIV. P. 65(c).

324. See CHARLES ALAN WRIGHT ET AL., 11A FEDERAL PRACTICE & PROCEDURE § 2954 (2d ed. 2010) (“The purpose of . . . [Federal Rule of Civil Procedure Rule 65(c)] is to enable a restrained or enjoined party to secure indemnification for the costs, usually not including attorney’s fees, and pecuniary injury that may accrue during the period in which a wrongfully issued equitable order remains in effect.” (internal citations omitted)).

325. See Grynberg, *supra* note 24, at 901-02 (arguing that formalist courts are unlikely to expand or develop trademark defenses).

326. *eBay Inc. v. MercExchange, L.L.C.*, 547 U.S. 388, 391 (2006).

327. *Id.* at 391-92; see also 5 MCCARTHY, *supra* note 173, at § 30:1 (“The Supreme Court has said that when a federal statute provides that injunctive relief may be granted in accordance with the principles of equity, a plaintiff seeking a permanent injunction must satisfy the traditional equitable four-factor test before a court may grant such relief. In no case is a plaintiff automatically entitled to a permanent injunction.”).

disserved by a permanent injunction.³²⁸

Instead of assuming that irreparable harm is present and giving short shrift to the public interest factor, courts should be encouraged to weigh the public's interests and give them equal weight in comparison to the other factors relevant to permanent injunctions. The same public interest factors discussed above, including those favoring trademark owners and the public generally, should be expressly weighed by the court.³²⁹ The public's interests can be unpacked to include consumer, user, and competitor interests.³³⁰ Notably, at this point in the litigation, a court has determined that consumer confusion is likely, there is commercial use, and a defense that may take into account a public interest, such as fair use, does not apply. However, the court should still balance the public interests along with any showing of irreparable harm and an inadequate remedy at law, even with the prospect of likely consumer confusion.³³¹ Indeed, in cases where the public interest is great, the court may require evidence of actual confusion in the form of survey evidence.³³² That analysis should inform the court's decision whether to carefully craft the injunction using disclaimers (to mitigate the development of any consumer confusion) or issue broad injunctive relief.

Moreover, as Professor Grynberg argues, the current model of trademark infringement places the mark holder in the position of advocating for both the public's interest and its own—to protect its goodwill—against the defendant's interest.³³³ He asserts that this analysis results in courts tipping the balance toward enjoining trademark infringement in close cases; thus, it overprotects trademarks.³³⁴ He argues that courts should also recognize the interests of the consumers who are not confused³³⁵ and thereby balance out the equation.³³⁶ The present proposal would take into account the interests of the unconfused consumer, such as the interest in free expression and obtaining comparative advertising information, but it would do so at the preliminary and permanent

328. *eBay*, 547 U.S. at 391.

329. See *supra* Part II.A.2.

330. Notably, the court can also consider the alleged infringer's interests, which may overlap with user or consumer interests, in considering the balance of hardships between the parties.

331. Cf. *Cliffs Notes, Inc. v. Bantam Doubleday Dell Publ'g Grp., Inc.*, 886 F.2d 490, 495 (2d Cir. 1989) (“taking into account that somewhat more risk of confusion is to be tolerated when a trademark holder seeks to enjoin artistic expression such as a parody”).

332. At least one scholar has criticized judges for treating consumers as “astoundingly naive, stunningly gullible, and frankly stupid” in assessing likelihood of confusion. See Ann Bartow, *Likelihood of Confusion*, 41 SAN DIEGO L. REV. 721, 723 (2004). Survey evidence is welcome where the public interest is implicated.

333. See Michael Grynberg, *Trademark Litigation as Consumer Conflict*, 83 N.Y.U. L. REV. 60, 72-84 (2008).

334. *Id.* at 61-64.

335. *Id.*

336. *Id.*

injunction stage through the public interest prong.

B. Eliminating Enforcement Efforts as a Basis for Obtaining, Retaining, and Expanding Rights

Another way to curb the expansion of trademark rights to the detriment of values (such as free speech and competition) is to disallow policing of trademarks as evidence of secondary meaning and mark strength to determine infringement, genericism, and fame. For example, courts should not consider evidence of prior suits, cease and desist letters, or other policing efforts in determining the strength of a particular mark or whether it has secondary meaning. As discussed above, some courts have used this evidence to determine whether a mark has a high level of distinctiveness.³³⁷ Although this evidence could be circumstantial and indicate the strength of a mark (because it demonstrates that perhaps consumers have had less of an opportunity to associate a mark with a mark used with goods from another source), its probative value does not outweigh the harmful effects of chilling speech or impairing other important interests.

The courts' consideration of policing evidence provides an incentive for mark holders to send more cease and desist letters and file more suits to demonstrate that the mark is indeed a strong mark. By disallowing that evidence, courts will remove an incentive to use litigation to stifle potentially protected speech or deny consumers the benefits that trademarks provide in enhancing competition. Moreover, advertising and sales revenue could still be used to determine acquired distinctiveness, fame, and strength of the mark, but courts should carefully examine that evidence to determine how the mark is used in that advertising and the materials used in connection with the sales.³³⁸ However, because of the focus on consumer perception in trademark law, courts should still be able to consider evidence of third party usage of a mark to determine whether that mark has acquired distinctiveness, is strong, is generic, or is abandoned—but only third party usage in the product or service market or related markets of the mark owner.³³⁹

It is important to note that this prohibition would be solely on the policing of third party use for certain purposes. In other situations, this evidence is directly relevant—for example, to demonstrate that a party has not unreasonably delayed in enforcing its trademark rights so that a laches defense may not apply. Moreover, there is an additional concern with any rights arising from a failure to assert rights against a party, such as an implied license. In those situations, the evidence of policing activity for the purpose of demonstrating that a mark holder

337. See *supra* Part II.B.1.

338. This analysis is regularly performed in the trade dress context in examining the relevancy of advertising by only accounting for “look for” advertising in determining acquired distinctiveness. RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 16 (1995).

339. Courts should also consider third party usage in related and unrelated markets to determine fame.

has not delayed in enforcing its rights or granted rights to another is directly relevant and should be considered for those limited purposes.

These proposals can be viewed as complementary to other reform efforts such as Professors Lemley and McKenna's proposal to add a materiality requirement to the infringement inquiry in certain cases.³⁴⁰ This would narrow infringement to those cases in which a consumer would find the confusing use material in its decision to purchase for specific types of confusion.³⁴¹ The materiality element would be similar to the analytical requirements in determining whether a mark is deceptive or primarily geographically deceptively misdescriptive.³⁴² The proposals also complement Professor McGeeveran's revision of the fair use defenses, including: keeping or adopting by statute Judge Kozinski's commercial fair use defense to dilution claims; extending the news reporting and commentary exclusion to federal dilution actions to infringement; adopting a defense for artistic and fictional work; and creating a defense for trademark use in a political campaign.³⁴³

CONCLUSION

Trademark protection has expanded over the last seventy years. Courts, Congress, and commentators are now searching for a way to reconcile the rights of the relevant stakeholders. This Article highlights some of the problems with the policy and structure of trademark law that make it on one hand effective, but on the other hand ineffective in achieving its numerous goals. As previously described, the consumer search cost theory has not provided the best normative framework to anchor the development of trademark doctrine, and the rights of trademark owners are expanding. The first proposal in this Article attempts to ensure that other interests besides those of the mark owner are considered in awarding injunctive relief, while the second proposal attempts to rein in the mark

340. See Lemley & McKenna, *supra* note 99, at 445-46 ("The solution, in our view, is simple: not to categorically rule out cases involving those other forms of confusion, but to limit those claims so as to increase the benefits of those still actionable and decrease their costs. The easiest way to do so would be to import into trademark law the materiality requirement courts have created in the false advertising context and apply it in any case based on confusion that does not relate to source or control over quality."); see also McKenna, *Trademark Use*, *supra* note 14, at 825 (tentatively suggesting that a plaintiff "prove that the defendant's use is likely to cause confusion that will materially affect consumer purchasing decisions").

341. See Lemley & McKenna, *supra* note 99, at 445-46.

342. See *In re Cal. Innovations, Inc.*, 329 F.3d 1334, 1336 (Fed. Cir. 2003); *In re Budge Mfg. Co.*, 857 F.2d 773, 775 (Fed. Cir. 1988). For a discussion of geographic marks, see Robert Brauneis & Roger E. Schechter, *Geographic Trademarks and the Protection of Competitor Communication*, 96 TRADEMARK REP. 782 (2006) and Joseph C. Daniels, Note, *The Branding of America: The Rise of Geographic Trademarks and the Need for a Strong Fair Use Defense*, 94 IOWA L. REV. 1703 (2009).

343. See generally McGeeveran, *Fair Use*, *supra* note 121, at 115-21; William McGeeveran, *The Trademark Fair Use Reform Act*, 90 B.U. L. REV. 2267 (2010).

holder's efforts to create and expand rights through the enforcement of its marks. These proposals should limit some of the expansion of trademark law caused by its reliance on consumer perception dictated by the consumer search cost theory, and perhaps they will restore some balance to trademark law.

WRONGFULLY INCARCERATED, RANDOMLY COMPENSATED—HOW TO FUND WRONGFUL-CONVICTION COMPENSATION STATUTES

DEBORAH MOSTAGHEL*

INTRODUCTION

It is sadly true that there are people in this country who are sentenced to prison, and even death, for crimes they did not commit. Some have been exonerated and released, largely as the result of innocence projects that have helped prisoners assemble DNA evidence that shows they were not the perpetrators.¹ Some have been exonerated years after they died in prison. Many others are no doubt never exonerated.² For a wrongfully convicted person, exoneration is the end of one road but only the beginning of another. Unbelievably, exonerees starting out on the road back to society find that they get little to no help from the justice system. Offenders on parole and convicts who complete their sentences may receive more services than released persons who committed no crime at all.³ This Article is based on the premise that states must pay compensation to innocent persons who have suffered wrongful imprisonment. The Article explains why exoneration is not enough. It then discusses theoretical justifications that support the payment of compensation and refutes objections to making such payments. Finally, it lays out concrete ways in which states may budget for wrongful-conviction compensation statutes. Although monetary compensation can hardly make up for years of wrongful imprisonment, providing compensation is the least a state can do after an

* Associate Professor and Director of Legal Writing and Research, Golden Gate University School of Law. J.D., 1988, University of Utah College of Law.

1. For example, Gary Dotson's 1979 conviction for rape was vacated and the charges dismissed on August 14, 1989. Samuel R. Gross et al., *Exonerations in the United States 1989 Through 2003*, 95 J. CRIM. L. & CRIMINOLOGY 523, 523 (2005). This was the "first exoneration by DNA evidence . . . [to take] place in the United States." Lauren C. Boucher, Comment, *Advancing the Argument in Favor of State Compensation for the Erroneously Convicted and Wrongfully Incarcerated*, 56 CATH. U. L. REV. 1069, 1069 (2007). One hundred ninety-eight post-conviction exonerations based on DNA evidence occurred between 1989 and 2007. Abigail Penzell, *Apology in the Context of Wrongful Conviction: Why the System Should Say It's Sorry*, 9 CARDOZO J. CONFLICT RESOL. 145, 145 (2007) (citing *About Us: Mission Statement*, THE INNOCENCE PROJECT, <http://www.innocenceproject.org/about/Mission-Statement.php> (last visited Feb. 12, 2011)).

2. Unknown numbers of prisoners whose innocence was never established have died in prison. Looking at exonerations beginning in 1989, "[i]n four cases, states posthumously acknowledged the innocence of defendants who had already died in prison: Frank Lee Smith, exonerated in Florida in 2000; Louis Greco and Henry Tameleo, exonerated in Massachusetts in 2002; and John Jeffers, exonerated in Indiana in 2002." Gross et al., *supra* note 1, at 524.

3. Heather Weigand, *Rebuilding a Life: The Wrongfully Convicted and Exonerated*, 18 B.U. PUB. INT. L.J. 427, 429 (2009).

innocent person is exonerated.

Part I of this Article explains wrongful conviction terminology. Part II explains how the very same reasons that make it easy for innocent people to be convicted also make it difficult for them to be exonerated. Part III explains why governments must compensate the wrongfully incarcerated when they win their freedom. It discusses the reasons for and against government compensation, rejecting the argument that imprisoning and even executing innocent persons is a harsh but necessary cost of doing business. Part IV deals with the hard question of how to pay for wrongful-conviction compensation statutes. To ensure that wrongfully imprisoned persons actually obtain compensation, the Article makes concrete suggestions for funding compensation statutes. Wrongfully convicted persons who win their freedom deserve compensation statutes, and states can afford them.

I. TERMINOLOGY

Innocent people convicted and sentenced to prison are described as the “wrongly convicted,”⁴ the “wrongfully convicted,”⁵ the “erroneously convicted,”⁶ the “unjustly convicted,”⁷ and the “unjustly imprisoned,”⁸ among other terms.⁹ The prison sentence is also variably described as “wrongful imprisonment,”¹⁰ “wrongful conviction,”¹¹ or “unjust conviction.”¹² I make the stylistic choice of

4. Alberto B. Lopez, *\$10 and a Denim Jacket? A Model Statute for Compensating the Wrongly Convicted*, 36 GA. L. REV. 665 (2002).

5. John H. Blume, *The Dilemma of the Criminal Defendant with a Prior Record—Lessons from the Wrongfully Convicted*, 5 J. EMPIRICAL LEGAL STUD. 477 (2008); Shawn Armbrust, Note, *When Money Isn’t Enough: The Case for Holistic Compensation of the Wrongfully Convicted*, 41 AM. CRIM. L. REV. 157 (2004).

6. Boucher, *supra* note 1; Christine L. Zaremski, Comment, *The Compensation of Erroneously Convicted Individuals in Pennsylvania*, 43 DUQ. L. REV. 429 (2005).

7. Adele Bernhard, *Justice Still Fails: A Review of Recent Efforts to Compensate Individuals Who Have Been Unjustly Convicted and Later Exonerated*, 52 DRAKE L. REV. 703 (2004) [hereinafter Bernhard, *Justice Still Fails*]; Adele Bernhard, *When Justice Fails: Indemnification for Unjust Conviction*, 6 U. CHI. L. SCH. ROUNDTABLE 73 (1999) [hereinafter Bernhard, *When Justice Fails*]; Erin Ann O’Hara, *Victims and Prison [sic] Release: A Modest Proposal*, 19 FED. SENT’G REP. 130, 133 (2006) (using the term “innocent convicts”).

8. Shelley Fite, *Compensation for the Unjustly Imprisoned: A Model for Reform in Wisconsin*, 2005 WIS. L. REV. 1181.

9. Charles I. Lugosi, *Punishing the Factually Innocent: DNA, Habeas Corpus and Justice*, 12 GEO. MASON U. C.R. L.J. 233 (2002); Daniel S. Medwed, *The Zeal Deal: Prosecutorial Resistance to Post-Conviction Claims of Innocence*, 84 B.U. L. REV. 125 (2004).

10. Jonathan L. Entin, *Being the Government Means (Almost) Never Having to Say You’re Sorry: The Sam Sheppard Case and the Meaning of Wrongful Imprisonment*, 38 AKRON L. REV. 139 (2005).

11. Penzell, *supra* note 1; see also Keith A. Findley & Michael S. Scott, *The Multiple Dimensions of Tunnel Vision in Criminal Cases*, 2006 WIS. L. REV. 291, 291; Brandon L. Garrett,

the Anglo-Saxon “wrong/wrongful/wrongfully” over the Latinate “error/erroneous/erroneously” and the philosophical choice of “wrongful/wrongfully” over “unjust/unjustly.” “Wrongfully” describes how the convicting was done, whereas “unjustly” focuses on the result of the conviction. The injustice flows from the wrongfulness of the conviction.¹³ Those who win their freedom are commonly called “exonerees.”

II. BACKGROUND: WHY IT IS EASY TO BE CONVICTED AND DIFFICULT TO BE EXONERATED

From mythology to the Bible to Shakespeare to children’s tales, popular culture is replete with stories in which no one believes a speaker’s claim to be telling the truth.¹⁴ Modern American criminal law has its share of such stories. Fifty years ago a murder took place that resulted in one of the best known stories: the tale of “The Fugitive.” Told and retold on television and in the movies, the story focuses on Sam Sheppard, who was wrongfully accused of murdering his wife, convicted, and imprisoned for twelve years before being found not guilty.¹⁵ More recent examples include the teens accused of raping the “Central Park Jogger”¹⁶ and members of the Duke University lacrosse team accused of raping

Innocence, Harmless Error, and Federal Wrongful Conviction Law, 2005 WIS. L. REV. 35.

12. Bernhard, *Justice Still Fails*, *supra* note 7, at 715; Bernhard, *When Justice Fails*, *supra* note 7, at 73.

13. It is noteworthy that commentators do not use the word “mistakenly” to describe the convicting of innocent persons. The unspoken assumption seems to be that a mistake can be innocent, but a decision reached wrongfully is more blameworthy.

14. E.g., Greek mythological figures Laocoon (he warned the Trojans not to accept the Greeks’ gift of the Trojan Horse, but no one listened to him) and Cassandra (she could prophesy the future, but it was fated that no one believe her); Biblical figure Susanna (the elders who tried to seduce her knew no one would take her word over theirs); Shakespeare’s Othello (he refused to believe that Desdemona was faithful to him); the Little Red Hen (she warned the barnyard animals that she would not share her homemade bread if they did not help plant, harvest, and grind the wheat—they did not; she did not).

15. See Entin, *supra* note 10, at 139-41. Entin begins this article by reviewing a book written by the *Sheppard* prosecutor about the retrial of the *Sheppard* case twenty-five years after the original wrongful conviction. In Entin’s review, he highlights all the ways that the *Sheppard* decision was and remains troubling. He uses that starting point as a springboard to discuss new approaches to wrongful imprisonment.

16. A young female jogger was attacked, beaten, assaulted and left for dead in Central Park in April 1989, presumably by a group of teenage boys “willing” (participating in a violent spree) in the park. Five teens were convicted of the rape and other attacks that occurred in the park that night. In December 2002, as a result of a confession by the real attacker, the boys’ convictions were set aside. Lynnell Hancock, *The Press and the Central Park Jogger*, COLUM. JOURNALISM REV. 1, 1-2, Jan. 1, 2003, available at <http://www.4efren.com/resources/The+Press+and+the+Central+Park+Jogger.pdf>.

a woman hired to dance at a party.¹⁷ What makes us believe in a person's guilt when that person is not guilty?

A. Easy to Be Convicted

Innocent defendants are wrongfully convicted and imprisoned for a variety of reasons generally traceable to facets of the criminal justice system. These include "fallible eyewitness identification evidence and flawed eyewitness identification procedures, false confessions, jailhouse snitch testimony, police and prosecutorial misconduct, forensic science error or fraud, and inadequate defense counsel."¹⁸ Mistaken eyewitness testimony¹⁹ may result in a wrongful accusation. Crime lab error²⁰ or ineffective assistance of counsel²¹ may turn the wrongful accusation into a wrongful conviction. Even though innocent, an accused person may confess.²²

Moreover, incentives within the prosecutorial system may influence prosecutors to obtain convictions.²³ Individually, prosecutors with high conviction rates are more likely to advance on the job²⁴ and so may overzealously pursue easy targets. Institutionally, district attorneys who show high conviction rates may be able to garner more resources than those who seem to be less successful.²⁵

B. Difficult to Be Exonerated

The same circumstances that allow wrongful convictions, and the same incentives that promote them, make it as hard to be exonerated as it is easy to be convicted. For example, prosecutors may oppose post-conviction DNA testing sought by a convicted person.²⁶ Prosecutors may also oppose post-conviction

17. "Duke University and its men's lacrosse team came under national scrutiny after a Durham woman alleged she was assaulted at a March 2006 team party off-campus. On April 11, 2007, the North Carolina Attorney General's Office dropped all charges against three indicted team members, saying they are innocent" *News & Communications*, DUKE UNIV., http://www.dukenews.duke.edu/mmedia/features/lacrosse_incident/ (last visited Feb. 13, 2011).

18. Findley & Scott, *supra* note 11, at 292.

19. Boucher, *supra* note 1, at 1074; Lopez, *supra* note 4, at 675.

20. Boucher, *supra* note 1, at 1078; Lopez, *supra* note 4, at 677 (identifying eyewitness error and creating a mega-category of "police, prosecutor, and scientific (mis)conduct"); *see also* Garrett, *supra* note 11, at 95 n.302, 98 n.311, 99 n.312.

21. Boucher, *supra* note 1, at 1080.

22. Liliana Segura, *Why Would Someone Confess to a Crime He Did Not Do?*, CHI. SUN-TIMES, Oct. 7, 2007, at B1.

23. Medwed, *supra* note 9, at 134-37. For further discussion, see Findley & Scott, *supra* note 11, at 374. For discussion in terms of fair trial rights, see Garrett, *supra* note 11, at 69-110.

24. Medwed, *supra* note 9, at 135.

25. *Id.* at 134-37.

26. *Id.* at 127 n.10. Charles Lugosi also describes the case of *Godschalk v. Montgomery County District Attorney's Office*, 177 F. Supp. 2d 366 (E.D. Pa. 2001), in which defendant

evidentiary hearings sought by the defense to present new evidence such as confessions from actual perpetrators.²⁷ Even after a wrongfully convicted person is released, prosecutors strive to uphold convictions, often in spite of evidence that shows a defendant is innocent.

Various aspects of the institutional culture of prosecutors' offices contribute to the drive to uphold convictions.²⁸ Professionally, once a conviction has been attained, "both the individual prosecutor and the office may become vested in maintaining the integrity of the conviction."²⁹ Psychologically, neither individuals nor organizations want to admit mistakes.³⁰ Prosecutors, though, may develop some sort of personal stake in the outcome.³¹ Even prosecutors who realize that innocent people might sometimes be convicted may have an "'ends justifies the means' outlook" that keeps them from "acknowledging the worthiness of a post-conviction innocence claim."³²

Further, prosecutors view post-conviction innocence claims skeptically because there are so many of them.³³ In addition to the professional and psychological motivations, political incentives may influence prosecutors to

Godschalk was finally able to win post-conviction DNA testing. Even though the test results exonerated Godschalk, the district attorney claimed the test results were flawed instead of releasing Godschalk. When further DNA testing still exonerated Godschalk, the district attorney reluctantly recommended to the court of common pleas that he be released. The district attorney still, however, refused to believe that Godschalk was innocent. Lugosi, *supra* note 9, at 235 n.10 (citing Sara Rimer, *Convict's DNA Sways Labs, Not a Determined Prosecutor*, N.Y. TIMES, Feb. 6, 2002, at A14).

27. Medwed, *supra* note 9, at 128.

28. *Id.* at 132. The *Sheppard* case presents a prime example. Sam Sheppard was convicted of second degree murder of his wife in 1954 and was sentenced to life in prison. Entin, *supra* note 10, at 139. Twelve years later, in 1966, the Supreme Court granted habeas corpus review and set aside the conviction after holding that pretrial publicity had tainted the original trial. *Id.* Later that year, Sheppard was acquitted in a retrial. *Id.* He died in 1970, and in 1995, his estate sought a declaration of innocence against the state so that the estate could seek compensation for Sheppard's wrongful imprisonment. *Id.* at 139-40. In the civil trial to establish innocence, the jury ruled against the estate even though Sheppard's conviction had been set aside and he had been acquitted. *Id.* at 140. The prosecutor in the civil trial, William Mason, set out to prove that Sam Sheppard was guilty in spite of the evidence that led to acquittal in his retrial. Entin explains that Mason, in his 2003 book about the civil trial, "sought to vindicate the honor and reputation of his office and of the other law enforcement agencies that had handled the Sheppard case from the beginning." *Id.* at 149.

29. Medwed, *supra* note 9, at 136.

30. *Id.* at 138.

31. This seems to have been the case with the prosecutor in the 1995 civil trial at which Sam Sheppard's son sought to have Sam Sheppard declared innocent. See Entin, *supra* note 10, at 139.

32. Medwed, *supra* note 9, at 147.

33. *Id.* at 148. Under the "'needle in a haystack' view of innocence claims, efficiency considerations militate against prosecutors thoroughly reviewing all post-conviction motions." *Id.* at 149.

strive to uphold convictions in post-conviction innocence cases, just as these same incentives may have led them to prosecute the cases in the first place.³⁴ First, prosecutors who need to win elections may emphasize a tough-on-crime stance.³⁵ Once in office, they need a record of successful convictions to substantiate such claims. Second, and ironically, state legislation to compensate wrongly convicted individuals may actually motivate prosecutors to resist post-conviction innocence claims³⁶ because prosecutors may fear that they or their departments will end up paying for these claims.

Statistical studies support the observation that it is difficult to obtain an exoneration. In a study of exonerations in the United States between 1989 and 2003, Gross and colleagues found 340 documented exonerations.³⁷ This may sound like a large number; it is not. According to the authors,

We can't come close to estimating the number of false convictions that occur in the United States, but the accumulating mass of exonerations gives us a glimpse of what we're missing. . . . Almost all the individual exonerations that we know about are clustered in two crimes, rape and murder. They are surrounded by widening circles of categories of cases with false convictions that have not been detected: rape convictions that have not been reexamined with DNA evidence; robberies, for which DNA identification is useless; murder cases that are ignored because the defendants were not sentenced to death; assault and drug convictions that are forgotten entirely. Any plausible guess at the total number of miscarriages of justice in America in the last fifteen years must be in the thousands, perhaps tens of thousands.³⁸

Exonerations can be said to demonstrate a failure of our justice system.³⁹

34. See *id.* at 149. This is not always the case. For example, Nancy E. Ryan, trial division chief in the Manhattan district attorney's office, led the reinvestigation of the Central Park Jogger case. The report that she authored in 2002, concluding that the five young men in prison for the attack had been wrongfully convicted, generated controversy and criticism. "The report, which ruled that Matias Reyes was the lone attacker, was criticized by the police and by some in Ms. Ryan's office." John Eligon, *Turning Off the Phone After 20 Years on Call*, CITY ROOM BLOG-NYTIES.COM (Mar. 12, 2010 7:38 AM), <http://cityroom.blogs.nytimes.com/2010/03/12/turning-off-the-phone-after-20-years-on-call/>.

35. Medwed, *supra* note 9, at 127, 154.

36. *Id.* at 154.

37. Gross et al., *supra* note 1, at 524. Gross points out that his team looked only at exonerations that resulted from case-by-case investigations into the particular circumstances of exonerated persons. *Id.* at 523-24. Their study did not include mass exonerations or people wrongly convicted of crimes such as misdemeanor assault or routine felonies. *Id.* at 533-35.

38. *Id.* at 551; see also Lopez, *supra* note 4, at 671 (noting that "the extent of factually incorrect convictions in our system must be much greater than anyone wants to believe").

39. See Lopez, *supra* note 4, at 674. But see Justice Scalia's concurrence in *Kansas v. Marsh*, 548 U.S. 163 (2006), a case analyzing the constitutionality of a state statute that required the death penalty if the jury found beyond a reasonable doubt that certain conditions existed. The defense

Recently, there does seem to be a trend toward more exonerations,⁴⁰ yet the number of exonerations still remains relatively low. When, against the odds, a wrongfully convicted person is exonerated, the last thing he should have to contend with is a battle to obtain meaningful compensation. This is a further punishment that adds to the existing failures of our justice system.

III. GOVERNMENTS MUST COMPENSATE THE WRONGLY CONVICTED

A. *Exoneration Is Not Enough*

A wrongfully convicted and incarcerated person pins his hopes on being exonerated. But exoneration is not enough. Exonerating a convicted person does not mean that he or she is restored to the status quo ante. Convicted persons whose innocence is established may be able to walk out of prison, but it is not so easy to walk back into society. More often than not, they have been harmed in countless ways. They may have lost years of their lives, their families, the opportunity to go to school, or the chance to gain or keep employment.

In our rapidly changing world, employment skills exonerees possessed upon entering prison may be out of date by the time of their release. They probably face discrimination in future employment despite their exoneration. They have probably lost any savings they once had. They have probably lost confidence in their ability to direct and manage their lives. They may have gained a host of physical and psychological problems due to prison conditions; in fact, they may even have lower life expectancies as a result of their incarceration. Further, they have arguably lost their status and respectability in the eyes of society.

Yet upon being set free from prison, at the time when they are the most helpless, exonerated prisoners are released into the world with virtually nothing in hand. Exonerees may receive such token assistance as fifty dollars⁴¹ or “ten dollars and a denim jacket.”⁴² Beyond so-called “gate money,”⁴³ some states

had contended that the statute violated the Eighth Amendment. The majority rejected the argument. *Id.* at 172. The dissent argued that the Eighth Amendment’s application to a death penalty statute after 1989 should be assessed in light of the “repeated exonerations of convicts under death sentences, in numbers never imagined before the development of DNA tests.” *Id.* at 208 (Souter, J., dissenting). In a separate concurring opinion, Justice Scalia found fault with statistical studies referred to by the dissent. He said that the exonerations included in the statistics were not limited to cases of factual innocence. Further, he opined that it “it is utterly impossible to regard ‘exoneration’—however casually defined—as a failure of the capital justice system, rather than as a vindication of its effectiveness.” *Id.* at 194 (Scalia, J., concurring).

40. See Gross et al., *supra* note 1, at 523, 527-28.

41. Armbrust, *supra* note 5, at 158.

42. Lopez, *supra* note 4, at 669.

43. The Arizona “gate money” program is typical: discharged prisoners receive up to fifty dollars that they have accumulated through deductions from their wages in a prison work program or through gifts from families and friends. Prisoners who have not worked will be “gifted” fifty dollars. Inmates also receive a set of clothes and possibly a bus or train ticket “to the closest stop

provide no help at all to exonerees.⁴⁴ In all states, exonerees must enter the legal system again if they wish to gain anything beyond token assistance. It is a difficult fight that exonerees lose more often than not. Even in states that do provide compensation through statute, the trigger for compensation is still for the exoneree to file a lawsuit against the state.⁴⁵ Although exonerees in states with compensation statutes may seem to be better positioned to receive compensation, applying for compensation under these statutes is a difficult and expensive undertaking that may still leave exonerees empty-handed.

1. *Three Avenues of Legal Redress Offer Little Hope to Exonerees.*—Four avenues of legal redress exist, but all four are not available in all states. In states without compensation statutes, only three avenues may be available. First, an exoneree may try to get the state legislature to pass a private bill awarding compensation in his or her particular case. Second, an exoneree may be able to bring a federal civil rights lawsuit against the government. Third, the exoneree may be able to bring a state tort lawsuit against prosecutors, police, or defense lawyers. Very few exonerees succeed under these approaches.

a. *Private bills.*—The first avenue—the private bill—requires that “a wrongfully convicted person must lobby his state legislature to pass a private bill that dispenses money from the state treasury directly to the lobbying individual as a remedy for the injustice of being wrongfully convicted.”⁴⁶ Lobbying the state legislature means finding a state legislator who will sponsor the bill and enlist support for it “in both houses so that the bill will pass when it comes up for a vote. This can be a long and arduous process . . .”⁴⁷

Assuming a private bill passes, some awards are significant;⁴⁸ some are not. In *O’Neil v. State*, defendant Leonard O’Neil was convicted and sentenced to a term of ten to twenty-five years for armed robbery.⁴⁹ When the actual perpetrator confessed, O’Neil had already served three and a half years of his wrongful sentence. He sued for compensation, but the court of claims dismissed the suit

outside state lines.” KATE J. WILSON, CTR. FOR PUB. POLICY RESEARCH, STATE POLICIES AND PROCEDURES REGARDING “GATE MONEY” 1-2 (Oct. 2007), *available at* http://www.cdr.ca.gov/adult_research_branch/research_documents/gate_money_oct_2007.pdf.

44. States without wrongful imprisonment compensation statutes are Alaska, Arizona, Arkansas, Colorado, Delaware, Georgia, Hawaii, Idaho, Indiana, Kansas, Kentucky, Michigan, Minnesota, Nevada, New Mexico, North Dakota, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Washington, and Wyoming. *State Compensation, LIFE AFTER EXONERATION PROGRAM*, http://www.exonerated.org/content/index.php?option=com_content&view=article&id=52&itemid=55 (last visited Feb. 13, 2011).

45. See Boucher, *supra* note 1, at 1083, 1099.

46. Lopez, *supra* note 4, at 698.

47. Adele Bernhard, *A Short Overview of the Statutory Remedies for the Wrongly Convicted: What Works, What Doesn’t and Why*, 18 B.U. PUB. INT. L.J. 403, 408 (2009) (giving a full account of the procedure and discussing some of the awards made in various states via this route).

48. *Id.* (discussing Georgia exoneree Clarence Harrison and Alabama exoneree Freddie Lee Gaines, who were each awarded \$1 million).

49. *O’Neil v. State*, 469 N.E.2d 1010, 1012 (Ohio Ct. App. 1984).

for failure to state a claim. At this point, the Ohio General Assembly enacted special legislation authorizing him to file a claim for “loss of education and employment and general damages . . . from the erroneous imprisonment.”⁵⁰ It would seem that O’Neil could now be compensated, and he was; the court of claims awarded him \$6,967. O’Neil appealed this paltry amount. In reversing the judgment, the Ohio Supreme Court said that when a person is wrongfully convicted,

the legislature and legal system have a responsibility to admit the mistake and diligently attempt to make the person as whole as is possible where the person has been deprived of his freedom and forced to live with criminals. Indeed the legal system is capable of creating few errors that have a greater impact upon an individual than to incarcerate him when he has committed no crime.⁵¹

Viewed in that context, “[the court’s] award of \$6,967 for O’Neil’s three-and-one-half years of erroneous imprisonment is grossly inadequate, is against the manifest weight of the evidence, and shocks the conscience.”⁵² Although the Ohio Court of Appeals gave the court of claims the opportunity to rethink its shockingly low award in this case, in general, few private bills ever get passed.⁵³ Even when one does pass, it may not go very far toward curing the harm.

b. Federal civil rights lawsuits under § 1983.—The second possible avenue of redress is to bring a federal civil rights lawsuit against the municipality and the police. Such a suit is brought under 42 U.S.C. § 1983.⁵⁴ Section 1983 offers the plaintiff a chance to overcome the immunity of state actors. The theory is that the municipality or police deprived the plaintiff of a constitutional right while “acting under color of state law.”⁵⁵ The deprivation of a constitutional right removes the barrier of government immunity, allowing the plaintiff to sue units of government—such as municipalities and police departments—that are normally immune from suit. Section 1983 is the vehicle for “interpos[ing] the federal courts between the States and the people, as guardians of the people’s

50. *Id.*

51. *Id.* at 1013.

52. *Id.*

53. Lopez, *supra* note 4, at 699 (“Unfortunately, legislative compensation remains a longshot for most wrongly convicted individuals . . .”).

54. Section 1983 provides that

[e]very person who, under color of any statute, ordinance, regulation, custom, or usage, of any State or Territory or the District of Columbia, subjects, or causes to be subjected, any citizen of the United States or other person within the jurisdiction thereof to the deprivation of any rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured in an action at law, suit in equity, or other proper proceeding for redress

42 U.S.C. § 1983 (2006).

55. See Lopez, *supra* note 4, at 690-98.

federal rights.”⁵⁶

Before invoking these protections by filing a § 1983 lawsuit, the plaintiff must satisfy the requirements of 28 U.S.C. § 2513, the Unjust Conviction and Imprisonment Act. This is the federal government’s wrongful conviction statute,⁵⁷ and it contains two basic categories of requirements. First, subsection (a)(1) requires the plaintiff to use the court record or a court certificate to show that: the conviction was set aside or reversed; he was found not guilty in a new trial; or he was pardoned on the ground of innocence.⁵⁸ Next, under subsection (a)(2), the plaintiff must not have committed any of the acts with which he was charged, and his own misconduct must not have brought about the prosecution.⁵⁹

Once the requirements of 28 U.S.C. § 2513 are satisfied, the exoneree may continue his quest under § 1983. Since § 1983 claims center on problems in the criminal justice system that may lead to a deprivation of civil rights, § 1983 seems the ideal path for exonerees to redress civil rights violations that may have occurred in their state cases. Focusing narrowly on what plaintiffs must do to initiate § 1983 lawsuits suggests that these suits will be very difficult for many plaintiffs to bring. Often, a potential § 1983 plaintiff may not be able to satisfy the threshold requirements.

Brandon Garrett, who has represented exonerees in multiple wrongful conviction cases,⁶⁰ points out that a federal wrongful conviction suit may only be brought if the exoneree can show that his case terminated with “[v]acatur of the conviction” and that the cause of the wrongful conviction was official misconduct.⁶¹ The vacatur requirement that “a plaintiff can file a federal case challenging unconstitutional conduct resulting in a conviction only after that

56. *Mitchum v. Foster*, 407 U.S. 225, 242 (1972).

57. The Unjust Conviction and Imprisonment Act was passed in 1938. Although it may seem that the federal government was aware of wrongful conviction issues rather early, the first attempt at wrongful conviction compensation actually dates from a bill introduced in the Senate in 1912. S. DOC. NO. 974, 62ND CONG. 3D SESS. (1912).

58. 28 U.S.C. § 2513(a)(1) (2006) provides that

[a]ny person suing under section 1495 of this title must allege and prove that: (1) His conviction has been reversed or set aside on the ground that he is not guilty of the offense of which he was convicted, or on new trial or rehearing he was found not guilty of such offense, as appears from the record or certificate of the court setting aside or reversing such conviction, or that he has been pardoned upon the stated ground of innocence and unjust conviction

“Section 1495” refers to 28 U.S.C. § 1495—damages for unjust conviction and imprisonment—the text of which is, “The United States Court of Federal Claims shall have jurisdiction to render judgment upon any claim for damages by any person unjustly convicted of an offense against the United States and imprisoned.”

59. *Id.* § 2513(a)(2). These are basically the same requirements for bringing a claim under state compensation statutes.

60. See generally Garrett, *supra* note 11. For reference to cases Garrett handled, see *id.* at 43 n.30.

61. *Id.* at 53-54.

conviction is either vacated or pardoned”⁶² affects potential § 1983 plaintiffs in one of two ways. First, a wrongfully incarcerated prisoner cannot bring a § 1983 lawsuit while he pursues a state remedy. Instead, he must wait until his state causes of action have concluded. This adds years to the time he spends seeking redress. Second, and much more damaging, the wrongfully incarcerated prisoner who is not able to bring a state suit has no other recourse. If he cannot succeed in bringing and winning a state wrongful conviction claim, the state court will not set aside the conviction, and the door to a federal civil rights lawsuit will remain shut.

Assuming, however, that the § 1983 plaintiff can establish vacatur, he must next show that his injury resulted from official misconduct on the part of police or prosecutors.⁶³ Official misconduct may result from official acts, such as suppression of exculpatory evidence, suggestive eyewitness identification procedures, coerced confessions, or fabrication of evidence.⁶⁴ Misconduct may be at the root of such civil rights causes of action as malicious prosecution and retaliatory prosecution.⁶⁵ Whatever the source, it may be difficult to establish that misconduct occurred. For example, police behavior can be justified under a probable cause standard. According to Michael Avery, “probable cause[] is a low standard, ordinarily not difficult for law enforcement to meet.”⁶⁶ The difficulties in establishing vacatur and official misconduct, from the narrow point of view of the wrongfully incarcerated individual, suggest that a § 1983 remedy may be remote.

A broader focus, however, indicates a greater significance for § 1983 lawsuits; these suits are important more as vehicles for exposing and addressing systemic deficiencies that led to the wrongful conviction rather than as means of redressing injuries to particular exonerees caused by the wrongful conviction. According to Garrett, although the two threshold requirements of vacatur and official misconduct result in the bringing of fewer wrongful conviction suits than “run-of-the-mill civil rights actions, the lawsuits that can be maintained involve the most egregious miscarriages of justice in which a conviction was vacated. Thus, through that filter, the cases brought may disproportionately involve misconduct implicating systemic failures.”⁶⁷ In discussing criminal justice reforms being both considered and adopted, Garrett seems to be saying that § 1983 suits are useful in a societal sense because they bring attention to the problems in criminal procedure that set defendants up for wrongful convictions.⁶⁸

The usual result of a criminal appeal is for the court to respond in favor of

62. *Id.*; see also *Heck v. Humphrey*, 512 U.S. 477, 486-87 (1994).

63. See Garrett, *supra* note 11, at 54.

64. *Id.* (citing *Brady v. Maryland*, 373 U.S. 83 (1963)). *Brady* will be discussed further in the rest of this section.

65. Michael Avery, *Obstacles to Litigating Civil Claims for Wrongful Conviction: An Overview*, 18 B.U. PUB. INT. L.J. 439, 439 (2009).

66. *Id.* at 442.

67. Garrett, *supra* note 11, at 54.

68. See *id.* at 99-102.

the criminal justice system rather than in favor of a criminal defendant, who may have been the victim of a suggestive lineup or coerced into confessing.⁶⁹ For example, in *Arizona v. Youngblood*, the Supreme Court held that potentially exculpatory evidence could be destroyed if it was not destroyed in bad faith.⁷⁰ But critics of the courts' tendency to favor law enforcement seem less concerned with the § 1983 plaintiffs—the exonerees—than with what § 1983 suits may accomplish. Civil suits under § 1983 generally focus on such “systemic deficiencies” as municipalities’ failures to train or supervise police, for example.⁷¹ Bringing attention to system-wide problems should lead to reform; indeed, “[t]he advance deterrent effect of such systemic claims will place the focus on what institutions can do to prevent wrongful convictions.”⁷² Commentators point out a connection between the growing numbers of § 1983 lawsuits and changes in police techniques that seem to signal institutional reform.⁷³

Certainly some exonerees do win § 1983 lawsuits,⁷⁴ and exonerees should continue to bring § 1983 suits. But they should be aware that their overall chance of recovery is slight.

c. *Common law tort suits*.—The third avenue of redress is common law tort suits. These may be brought in state court or in federal court, depending on the specific cause of action. Such suits are also difficult to win. A plaintiff may bring a state court tort suit against police or prosecutors based on malicious prosecution, but it is generally difficult to prove malice.⁷⁵ Relying on *Brady* requirements designed to guarantee a fair trial, a plaintiff may make a *Brady* due process claim alleging that the prosecutor’s failure to disclose material evidence prejudiced the plaintiff’s ability to mount an effective defense.⁷⁶ The difficulty here, according to Michael Avery, is that

[t]he Supreme Court has never resolved whether a criminal defendant’s due process right to obtain exculpatory evidence in the hands of the state, protected by *Brady v. Maryland* and its progeny, is bottomed on substantive or procedural due process. As a consequence, the lower

69. See *id.* at 58-62.

70. *Arizona v. Youngblood*, 488 U.S. 51, 57-58 (1988). This seems to be an example of moral hazard, discussed *infra* Part III.C.3.

71. Garrett, *supra* note 11, at 106-07.

72. *Id.* at 107.

73. See, e.g., *id.* at 45-46.

74. E.g., *Pierce v. Gilchrist*, 359 F.3d 1279 (10th Cir. 2004); *Newsome v. McCabe*, 319 F.3d 301 (7th Cir. 2003); *Jones v. City of Chi.*, 856 F.2d 985 (7th Cir. 1988) (suppressing exculpatory evidence); *People v. LeGrand*, 867 N.E.2d 374 (N.Y. 2007) (denying defense request to call an expert on causes of mistaken identification).

75. Garrett, *supra* note 11, at 50-51.

76. *Anderson v. State*, 196 S.W.3d 28, 36-37 (Mo. 2006). For a very clear explanation of how *Brady* violations may be established and used as the basis for a civil rights lawsuit, see Bernhard, *Justice Still Fails*, *supra* note 7, at 726-28.

federal courts are in disagreement on the issue.⁷⁷

Predicting outcomes of these cases is difficult.

A plaintiff may also bring a tort suit against his own defense attorney for ineffective assistance of counsel.⁷⁸ However, plaintiffs have lost these suits even when their defense counsel slept through the trial,⁷⁹ presented no evidence during the sentencing phase of the trial,⁸⁰ or was intoxicated during the trial.⁸¹

Finally, plaintiffs may bring common law tort suits against the United States in federal court under the Federal Tort Claims Act (FTCA).⁸² The FTCA “evinces a waiver of sovereign immunity with respect to certain categories of torts committed by federal employees in the scope of their employment.”⁸³ These “certain categories” are all negligence-based causes of action because the Federal Tort Claims Act specifically excludes the intentional torts of “assault, battery, false imprisonment, false arrest, malicious process, abuse of process, libel, slander, misrepresentation, deceit, [and] interference with contract rights.”⁸⁴ The FTCA provides that the United States may only be found liable “under circumstances where the United States, if a private person, would be liable to the claimant in accordance with the law of the place where the act or omission occurred.”⁸⁵ This means that even though the plaintiff is suing the federal government in federal court, the law governing the cause of action will be that of the state “in which the allegedly tortious acts or omissions occurred.”⁸⁶ This law-of-the-state requirement is how the federal government establishes jurisdiction over the cause of action. A federal court will not have jurisdiction to hear an FTCA case unless a state court would have had jurisdiction if the tortfeasor had been a private actor. Claims under the FTCA can only be brought

77. Avery, *supra* note 65, at 446 (internal citation omitted).

78. Bernhard, *Justice Still Fails*, *supra* note 7, at 736-37.

79. *United States v. Petersen*, 777 F.2d 482, 484 (9th Cir. 1985). The plaintiff was unable to establish ineffectiveness of counsel because defense counsel had not slept through “a substantial portion of the trial.” *Id.* By comparison, in *Burdine v. Johnson*, the Fifth Circuit Court of Appeals affirmed the district court’s conclusion that “sleeping counsel is equivalent to no counsel” when defense counsel “repeatedly dozed or slept as the State questioned witnesses” during Burdine’s murder trial. *Burdine v. Johnson*, 262 F.3d 336, 339 (5th Cir. 2001); see also *McFarland v. Texas*, 928 S.W.2d 482, 505 (Tex. Crim. App. 1996) (en banc), abrogated by *Mosley v. State*, 983 S.W.2d 249 (Tex. Crim. App. 1998).

80. *Mitchell v. Kemp*, 762 F.2d 886, 888 (11th Cir. 1985).

81. *Haney v. State*, 603 So. 2d 368, 377-78 (Ala. Crim. App. 1991).

82. The FTCA is codified at 28 U.S.C. §§ 1346(b), 2671-80 (2006). 28 U.S.C. § 1346(b) is recognized as preempted by *Moore v. Potter*, 605 F. Supp. 2d 731, 734 (E.D. Va. 2009) (preempted for postal workers by the Postal Reorganization Act).

83. *Bolduc v. United States*, 402 F.3d 50, 55 (1st Cir. 2005).

84. 28 U.S.C. § 2680(h).

85. *Id.* § 1346(b)(1).

86. *Bolduc*, 402 F.3d at 56.

if the “challenged government conduct” has a “parallel in the private sector.”⁸⁷ Further, the behavior must not fall under the discretionary exception to the FTCA,⁸⁸ which allows some government conduct to be sheltered from tort liability.⁸⁹ This exception creates difficulty for wrongful-conviction plaintiffs because even though a federal official’s actions may have been negligent, the official will be protected from suit when the negligent acts resulted from policy decisions the official was required to make in his official role.

The case of *Bolduc v. United States* illustrates both of these concepts.⁹⁰ In *Bolduc*, two men were wrongly convicted and incarcerated for a robbery in Wisconsin.⁹¹ They spent eight years in prison before the actual perpetrator was identified.⁹² After their release, they sued the United States for the negligence of an agent who withheld evidence that would have exonerated them and for the negligent supervision of the agent by the agent’s supervisor.⁹³ In a bench trial, the court denied the claim, and the claimants appealed.⁹⁴ In affirming the district court opinion, the First Circuit Court of Appeals first analyzed the negligence claim.⁹⁵ The court applied Wisconsin negligence law⁹⁶ and concluded that the claim failed under Wisconsin law “[b]ecause Wisconsin’s recognition of a governmental duty to disclose exculpatory evidence does not ground private liability under that state’s law.”⁹⁷ That is, Wisconsin could recognize a governmental duty to disclose exculpatory information. But to prevail, the plaintiffs needed to show—and could not—that Wisconsin

imposed *private* liability on a prosecutor or other state agent for a failure to disclose exculpatory evidence. That is a fatal flaw, for the federal government does not yield its immunity with respect to obligations that

87. See *id.* (citing *Sea Air Shuttle Corp. v. United States*, 112 F.3d 532, 536-37 (1st Cir. 1997)).

88. The FTCA’s discretionary exception section, codified at 28 U.S.C. § 2680(a), states that [t]he provisions of this chapter and section 1346(b) of this title shall not apply to . . . [a]ny claim based upon an act or omission of an employee of the Government, exercising due care, in the execution of a statute or regulation, whether or not such statute or regulation be valid, or based upon the exercise or performance or the failure to exercise or perform a discretionary function or duty on the part of a federal agency or an employee of the Government, whether or not the discretion involved be abused.

There are other exceptions, but this one is commonly encountered by wrongful-conviction plaintiffs.

89. *Bolduc*, 402 F.3d at 60.

90. See *id.* at 55-56.

91. *Id.* at 53.

92. *Id.* at 54.

93. *Id.*

94. *Id.*

95. *Id.* at 56.

96. *Id.*

97. *Id.* at 57.

are peculiar to governments or official-capacity state actors and which have no private counterpart in state law.⁹⁸

Next, the court analyzed the claim of negligent supervision, which failed because the FBI's supervision of the agent fell under the discretionary function exception to the FTCA.⁹⁹ The court reasoned that government agencies need to be able to carry out certain activities without constant threat of suit.¹⁰⁰ A court reviews "whether the conduct itself is discretionary, that is, 'a matter of choice for the acting employee.'"¹⁰¹ If so, the court then looks to see if that exercise of discretion involves policymaking because "[o]nly if the conduct is both discretionary and policy-driven will section 2680(a) strip the court of subject matter jurisdiction."¹⁰² The agent's supervisors' actions in this case were both discretionary and policy-related. Thus, they were protected because "this court has recognized, in the context of supervision, that in the absence of a statutory or regulatory regime that sets out particulars as to how an agency must fulfill its mandate, the development and management of a supervisory model is a matter of agency discretion."¹⁰³ As a result, the United States could not be liable.

As *Bolduc* illustrates, it is very difficult for wrongful-conviction plaintiffs to show that an official's actions would subject the official to liability if he were a private person. If plaintiffs succeed in making this link, more likely than not, the official's actions will be protected under the discretionary function exception.

In sum, private bills afford relief to few. Litigation is slow, costly, and uncertain. Adele Bernhard, who has championed compensation for the wrongfully convicted since 1999,¹⁰⁴ notes that "litigation has yielded mixed results. Few exonerated individuals have been compensated. And . . . others, no more deserving, have received enormous awards."¹⁰⁵ But until all states have fair and easily accessible compensation statutes, litigation—whether § 1983 federal suits or state common law tort suits—remains an essential tool for exonerees.¹⁰⁶

2. The Fourth Avenue, State Compensation Statutes, Holds the Most Promise, But Overly Restrictive Requirements Create Hardships for Exonerees.—The fourth avenue of redress for the wrongfully incarcerated is to sue for compensation under a state wrongful-conviction compensation statute. State compensation statutes offer the promise of speedier and more equitable damage awards. It is the premise of this Article that states without compensation statutes should pass them. This section comments on two features of some existing

98. *Id.*

99. *Id.* at 62.

100. *Id.* at 60 (citing *United States v. Varig Airlines*, 467 U.S. 797, 808 (1984)).

101. *Id.* (quoting *Berkovitz v. United States*, 486 U.S. 531, 536 (1988)).

102. *Id.*

103. *Id.* at 61.

104. See generally Bernhard, *When Justice Fails*, *supra* note 7.

105. Bernhard, *Justice Still Fails*, *supra* note 7, at 707 (citations omitted).

106. Indeed, one of Bernhard's goals is to "encourage more litigation" using new theories based on creative strategies. *Id.* at 726-27.

statutes that are challenging for exonerees: (1) actual innocence; and (2) possible reductions in awards under compensation statutes should the exoneree also win a § 1983 suit.

a. *Actual innocence.*—Compensation statutes, also called enabling statutes,¹⁰⁷ enable the exoneree to sue the state itself on the ground that the exoneree was wrongfully incarcerated and should be compensated. Twenty-five states, the District of Columbia, and the federal government have passed enabling statutes.¹⁰⁸ The exoneree must meet various requirements, depending on the state.¹⁰⁹ Under all the statutes, an exoneree must demonstrate his innocence of the crime for which he was convicted.

Wrongful-conviction compensation statutes generally have four requirements. First, the exoneree must have been convicted of a crime—typically, a felony.¹¹⁰ Second, under most statutes, the claimant must not have pleaded guilty to the charged offense.¹¹¹ Third, the claimant must have been sentenced to incarceration and have actually served time as a result of the conviction.¹¹² Fourth, the claimant must establish actual—also called factual—innocence.

The goal of the fourth requirement is to ensure that the claimant is actually innocent. Establishing factual innocence calls for “an official act declaring a defendant not guilty of a crime for which he or she had previously been

107. See Zaremski, *supra* note 6, at 436.

108. Bernhard, *supra* note 47, at 409 n.49 (citing *Compensating the Wrongly Convicted Fact Sheet*, THE INNOCENCE PROJECT, http://www.innocenceproject.org/Content/Compensating_The_Wrongly_Convicted.php (last visited Feb. 13, 2011)).

109. New Hampshire is one state whose statute does not provide the elements of a *prima facie* case. The exoneree may choose an administrative remedy or a judicial remedy, under which the exoneree brings an action against the state “for time unjustly served in the state prison when a person is found to be innocent of the crime for which he was convicted.” N.H. REV. STAT. ANN. § 541-B:14 (2007 & Supp. 2009). Maryland requires that an exoneree receive a full pardon. The pardon must state that the conviction was conclusively demonstrated to have been in error. MD. CODE ANN., STATE FIN. & PROC. § 10-501(b) (West, Westlaw through 2010 Reg. Sess.).

110. E.g., ALA. CODE § 29-2-156 (2003); CAL. PENAL CODE § 4900 (2010). However, New York permits claims based on misdemeanor convictions. N.Y. CT. CL. ACT § 8-b.2 (McKinney 2003 & Supp. 2011). An Illinois court has interpreted the “crime” requirement of 705 ILL. COMP. STAT. ANN. 505/8(c) (West, Westlaw through 2010 Reg. Sess.) to be satisfied by any offense for which an exoneree suffered wrongful imprisonment. *Smith v. State*, 26 Ill. Ct. Cl. 290, 298 (West 1969).

111. For example, the Iowa statute defines a “wrongfully imprisoned person” as one who “did not plead guilty to the public offense charged, or to any lesser included offense, but was convicted by the court or by a jury of an offense classified as an aggravated misdemeanor or felony.” IOWA CODE ANN. § 663A.1(1)(b) (West, Westlaw through 2010 Reg. Sess.). An Oklahoma statute requires that the exoneree be charged with the commission of a felony and that he not have pleaded guilty. OKLA. STAT. ANN. tit. 51, § 154(B)(1) (West, Westlaw through 2010 2d Reg. Sess.).

112. E.g., N.C. GEN. STAT. § 148-82 (LEXIS through 2010 Reg. Sess.); N.Y. CT. CL. ACT § 8-b.3 (McKinney 2003 & Supp. 2011); IOWA CODE ANN. § 663A.1.

convicted.”¹¹³ In 2005, Samuel Gross and colleagues published results of a study of the 340 exonerations that took place in the United States between 1989 and 2003.¹¹⁴ Gross identified four specific types of official acts by which the defendants in the study were exonerated. These official acts were gubernatorial pardons, court dismissal of charges, acquittal at retrial, and posthumous state acknowledgment of innocence.¹¹⁵ However, not all state statutes recognize all four forms. In five states, the only acceptable official act to establish factual innocence is a pardon from the governor, regardless of any other proof of innocence that exists.¹¹⁶ This stringent requirement can lead to harsh results. To illustrate, consider the 340 exonerations studied by Gross. In the exonerations of these 340 people, governors issued forty-two pardons.¹¹⁷ Under all existing compensation statutes, those forty-two exonerees who were lucky enough to receive pardons would be eligible to sue or otherwise apply for compensation. But under the statutes of the five states requiring pardons, the remaining exonerees would not have been eligible for compensation because they were not pardoned by governors.¹¹⁸ Courts dismissed criminal charges in the cases of 263 of the 340 exonerees.¹¹⁹ Thirty-one exonerees were acquitted at retrials.¹²⁰ Four exonerees died in prison; their innocence was posthumously acknowledged.¹²¹ What this means is that under the compensation statutes requiring pardons, 298 people out of the 340 who were exonerated—298 people officially declared not guilty of a crime—would receive no compensation because a governor did not give them a pardon. Such a stringent requirement makes it unnecessarily difficult for wrongfully incarcerated persons to establish innocence.

In addition to the four official acts establishing factual innocence that Gross

113. Gross et al., *supra* note 1, at 524.

114. *Id.* at 523-24.

115. *Id.* Texas posthumously exonerated Timothy Cole in 2009, ten years after his death in prison. He received a posthumous pardon on March 1, 2010. Cole, a college student, was convicted in 1985 of raping a Texas Tech student and sentenced to twenty-five years in prison. He was cleared by DNA evidence in 2008, exonerated in 2009, and pardoned in 2010. Texas Governor Rick Perry granted him a full pardon on Monday, March 1, 2010, as soon as he received a legal opinion from the Texas attorney general saying that the governor had such authority. *See Timothy Cole: A Tragic Story Begets Hope for the Future*, INNOCENCE PROJECT OF TEX., <http://www.ipoftexas.org/index.php?action=timothy-cole> (last visited Feb. 14, 2011).

116. Boucher, *supra* note 1, at 1085 n.113 (listing the five states that require pardons: California, Illinois, Maine, Maryland, and North Carolina).

117. Gross et al., *supra* note 1, at 524.

118. CAL. PENAL CODE § 4900 (2010); 705 ILL. COMP. STAT. ANN. 505/8(c) (West, Westlaw through 2010 Reg. Sess.); ME. REV. STAT. ANN. tit. 14, § 8241(2) (West, Westlaw through 2009 2d Reg. Sess.); MD. CODE ANN., STATE FIN. & PROC. § 10-501 (West, Westlaw through 2010 Reg. Sess.); *see also* N.C. CONST. art. III, § 5(6); N.C. GEN. STAT. § 148-82 (LEXIS through 2010 Reg. Sess.); 88 Md. Op. Att'y Gen. 03-007 (2003).

119. Gross et al., *supra* note 1, at 524.

120. *Id.*

121. *Id.*

described in 2003, some statutes require yet another official act before the exoneree may be considered for compensation. Namely, the claimant must produce proof from the trial court that he was innocent of the crime.¹²² Ohio does not require a gubernatorial pardon, but it does require an express finding of innocence. The claimant must secure a “determin[ation] by a court of common pleas that the offense of which the individual was found guilty . . . either was not committed by the individual or was not committed by any person.”¹²³ Maine requires not only that the exoneree receive a full pardon,¹²⁴ but also that the exoneree be found innocent by a court.¹²⁵ However, not all states require such restrictive proof of innocence as Maine does. Alabama, for example, does not require an express finding of innocence. The claimant must demonstrate that the conviction was vacated or reversed and that the accusatory instrument was dismissed either on the ground of innocence or a ground “consistent with innocence.”¹²⁶ By comparison, New Jersey requires that the claimant show that he “did not commit” the crime.¹²⁷

Atypically, New York’s scheme allows for a claim based not only on pardons and factual innocence, but also on legal innocence.¹²⁸ To succeed on a claim of legal innocence, a claimant must show that the alleged facts did not constitute a felony or misdemeanor against the state.¹²⁹ This is a less restrictive standard; most statutes require that the exoneree not have committed acts that constitute a felony.¹³⁰ Although committing acts that do not constitute a felony seems to correspond to not committing acts that do constitute a felony, it might be easier to demonstrate that actions did not equal a felony than to demonstrate that no actions were taken.

Inequalities exist in the kind and amount of compensation as well as in the standards for establishing innocence. Some statutes determine an exoneree’s compensation on a case-by-case basis.¹³¹ Some award a set amount regardless

122. *Haddad v. Dep’t of Rehab. & Corr.*, No. 01AP-1130, 2002 WL 1163917, at *2-3 (Ohio Ct. App. June 4, 2002). Although unpublished, *Haddad* has been cited in a later—also unpublished—decision. *See Mickey v. Ohio Dep’t of Rehab. & Corr.*, No. 02AP-539, 2003 WL 116152, at *4 (Ohio Ct. App. Jan. 14, 2003).

123. OHIO REV. CODE ANN. § 2743.48(A)(5) (West, Westlaw through 2010 legislation).

124. ME. REV. STAT. ANN. tit. 14, § 8241(2)(c) (West, Westlaw through 2009 2d Reg. Sess.).

125. *Id.* § 8241(2)(d). At the time of the Gross study, Maine had not exonerated anyone.

126. ALA. CODE § 29-2-157(2) (2003).

127. N.J. STAT. ANN. § 52:4C-2 (2011).

128. N.Y. CT. CL. ACT § 8-b.3 (McKinney 2003 & Supp. 2011).

129. N.Y. CRIM. PROC. LAW § 440.10(1) (McKinney 2005) sets forth the grounds for setting aside convictions; *see also id.* § 470.20(1)-(3), (5) (McKinney 2009).

130. *E.g.*, ALA. CODE § 29-2-161(e); CAL. PENAL CODE § 4900 (2010); N.J. STAT. ANN. § 52:4C-3(b); OHIO REV. CODE ANN. § 2743.48(A)(5) (West, Westlaw through 2010 legislation); W. VA. CODE ANN. § 14-2-13a(f)(4) (West, Westlaw through 2011 Reg. Sess.).

131. *E.g.*, N.Y. CT. CL. ACT § 8-b.6; TENN. CODE ANN. § 9-8-108(a)(7)(A) (West, Westlaw through 2010 Reg. Sess.) (lifetime limit of \$1 million); W. VA. CODE ANN. § 14-2-13a(g).

of the length of time the exoneree was in prison.¹³² Others award a set amount based on the length of the wrongful incarceration.¹³³

One consistent aspect of compensation statutes is that innocence will not be considered established if reversal occurred merely because of procedural or jurisdictional errors. In *Walden v. State*, the Ohio Supreme Court described a defendant's acquittal in a criminal trial as "a determination that the state has not met its burden of proof on the essential elements of the crime. It is not necessarily a finding that the accused is innocent."¹³⁴ Claimants are consistently required to establish that the conviction was reversed on relevant grounds.¹³⁵ Relevant grounds do not include the following: claims of double jeopardy,¹³⁶ "not guilty by reason of insanity;"¹³⁷ a court's failure to comply with provisions of the criminal statute;¹³⁸ the fact that the statute under which a claimant was convicted was later found unconstitutional,¹³⁹ or that the convicting court lacked jurisdiction.

Though focusing on the jurisdictional issue, the case of *Osborn v. United States*¹⁴⁰ illustrates the requirement that the claimant must demonstrate either that he did not commit the act or that the act he committed did not constitute a crime. In *Osborn*, four servicemen already court-martialed and in prison for other offenses were charged with premeditated murder of a fellow inmate under the 92nd Article of War. After being found guilty, one of the co-defendants filed a habeas corpus petition on the ground that under the 92nd Article of War, courts-martial did not have jurisdiction over trials for rape or murder during peacetime.¹⁴¹ The court granted the petition because the nation had been at peace when the defendants were charged and court-martialed.¹⁴² Co-defendant Osborn then filed his own habeas motion, the granting of which resulted in his discharge from prison.¹⁴³ Then Osborn sued the United States for wrongful

132. E.g., N.H. REV. STAT. § 541-B:14 (2007 & Supp. 2010).

133. E.g., ALA. CODE § 29-2-159(a); CAL. PENAL CODE § 4904; 705 ILL. COMP. STAT. ANN. 505/8(c) (West, Westlaw through 2010 Reg. Sess.) (amount varies with length of time incarcerated); TEX. CIV. PRAC. & REM. CODE ANN. § 103.052(a)-(b) (West, Westlaw through 2009 Reg. Leg.).

134. *Walden v. State*, 547 N.E.2d 962, 966 (Ohio 1989).

135. See, e.g., D.C. CODE § 2-422(1) (2011).

136. *Fudger v. State*, 520 N.Y.S.2d 950, 952-53 (N.Y. App. Div. 1987).

137. See generally *Ebberts v. State Bd. of Control*, 148 Cal. Rptr. 543 (Cal. Ct. App. 1978).

138. *Mickey v. Ohio Dep't of Rehab. & Corr.*, No. 02AP-539, 2003 WL 116152, at *4 (Ohio Ct. App. Jan. 14, 2003) (holding that lower court failed to comply with § 2967.28(B)(3) of the Ohio Code).

139. *Lambert v. State Claims Bd.*, No. 78-306, 1979 WL 30360, at *1-2 (Wis. Ct. App. 1979) (finding that the claimant had violated the statute when it was in effect).

140. 322 F.2d 835, 838 (5th Cir. 1963).

141. *Id.* at 837-38.

142. *Id.* at 838.

143. *Id.*

imprisonment.¹⁴⁴ The district court “found that the record showed that the plaintiff’s conviction had not been reversed or set aside upon the stated ground of innocence and unjust conviction, and that Osborn had failed to show that he had not committed the acts with which he was charged.”¹⁴⁵ The court of appeals affirmed.¹⁴⁶ At best, Osborn could demonstrate that he was discharged because the court-martial lacked jurisdiction. That showing would not satisfy the wrongful conviction statute; Osborn could not demonstrate that he did not commit the murder, and he also could not demonstrate that the act of murder did not constitute a crime.¹⁴⁷ In short, whether federal or state, wrongful-conviction compensation statutes are intended to “separate those who were wrongfully imprisoned from those who have merely avoided criminal liability.”¹⁴⁸

As important as it is for states to pass enabling legislation for exoneree compensation, it is equally important that they analyze the need for onerous requirements and widely varying calculation methods. Exonerees may be shut out of compensation because of needlessly strict requirements even in states that have passed compensation statutes. Relatively few exonerees succeed in establishing all the elements required under compensation statutes. Widely varying requirements for establishing innocence or determining compensation make recovery more difficult for some exonerees than for others merely based on where they live.¹⁴⁹ Thus, a goal of statutory revision should be to create consistency among statutes.

b. Possible reductions in compensation statute awards should the exoneree also win a § 1983 lawsuit.—State compensation statutes may require as a condition of compensation that the exoneree agree to give up the third avenue of legal redress,¹⁵⁰ which is to bring suit against the state on common law tort theories. This is a reasonable requirement because the point of the compensation statute is to streamline and make more certain the exoneree’s route to compensation.¹⁵¹

A New York case suggests that courts may treat § 1983 suits the same way by offsetting one award against another should an exoneree win a § 1983 suit and

144. *Id.*

145. *Id.*

146. *Id.* at 842-43.

147. *Id.* at 842.

148. *Walden v. State*, 547 N.E.2d 962, 967 (Ohio 1989).

149. An Iowa statute allows a claim to go forward after the exoneree obtains a district court order adjudging the exoneree to be a “wrongfully imprisoned person.” IOWA CODE ANN. § 663A.1 (West, Westlaw through 2010 Reg. Sess.). Other statutes (those of California, Illinois, Maine, Maryland, and North Carolina, noted *supra* note 118) require pardons from the governor.

150. See *supra* Part III.A.1.

151. In at least one instance, however, a court ruled that separate causes of action for the same wrongful imprisonment could go forward (under the state compensation statute and the state tort claims act) if negligence by a state employee was responsible for the wrongful imprisonment “based on acts or omissions apart from the process responsible for the conviction and wrongful imprisonment.” *Cox v. State*, 686 N.W.2d 209, 215 (Iowa 2004).

attempt to claim compensation under a state statute as well. In *Carter v. State*, an exoneree was compensated under § 1983 for wrongful conviction.¹⁵² He also brought suit under New York's Unjust Conviction and Imprisonment Act, which is part of the state's Court of Claims Act. The appellate court ruled that the exoneree's recovery under § 1983 barred recovery under the state's statutory compensation remedy.¹⁵³ The court reasoned that the success of the claimant's suit under § 1983 showed that he was not a person "frustrated in seeking legal redress," which was the class of persons the Unjust Conviction and Imprisonment Act sought to protect.¹⁵⁴ Further, the claimant could not recover because he had already been compensated for the injuries resulting from his wrongful conviction.¹⁵⁵ Thus, exonerees with state statutory remedies may choose to file § 1983 lawsuits anyway, but they should be aware of the potential for offsets in their statutory damage awards.

B. Theoretical Justifications Support the Payment of Compensation

Once a wrongfully convicted person has been found innocent, the state should pay compensation. Two theoretical justifications—the instrumental and corrective justice theories—support the payment by a state to an exoneree.

Compensation statutes arose in the context of providing aid to crime victims.¹⁵⁶ The goal of compensation statutes for crime victims is to "give financial or other aid to victims or their survivors."¹⁵⁷ Such statutes grew out of a "rights theory"¹⁵⁸ positing that a state that fails 'to protect its citizens from crime is obligated to provide compensation to those who become victims.'"¹⁵⁹ Funding these statutes, however, may complicate compensation for wrongly convicted persons. Statutes that provide compensation for crime victims do so through fines imposed on criminal wrongdoers.¹⁶⁰ In the case of wrongly convicted

152. *Carter v. State*, 546 N.Y.S.2d 648, 650 (N.Y. App. Div. 1989).

153. *Id.*

154. *Id.* (quoting N.Y. CT. CL. ACT § 8-b(1) (McKinney 1989)).

155. *Id.* However, in at least one case, an exoneree did have the right to pursue both a § 1983 remedy and a state statutory remedy. *See, e.g., State v. Oakley*, 181 S.W.3d 855, 857 (Tex. Crim. App. 2005) (granting rule 53.7(f) motion), *aff'd in part and rev'd on other grounds*, 227 S.W.3d 58 (Tex. 2007).

156. Federal crime victim compensation statutes have been in place in this country since the 1980s; similar state statutes have been in place since the 1990s. Deborah M. Mostaghel, *Wrong Place, Wrong Time, Unfair Treatment? Aid to Victims of Terrorist Attacks*, 40 BRANDEIS L.J. 83, 87 n.17 (2001).

157. *Id.* at 87. Compensation statutes provide aid to the victim; by contrast, under restitution statutes, victims or their survivors can seek restitution from the perpetrator.

158. This "rights theory" is the same as the corrective justice theory (*see infra* Part III.B.2); it supports compensation for victims of the criminal justice system no less than for victims of crime.

159. Mostaghel, *supra* note 156, at 87 (quoting Charlene L. Smith, *Victim Compensation: Hard Questions and Suggested Remedies*, 17 RUTGERS L.J. 51, 62 (1985)).

160. *Id.* at 87.

persons, there is no easily targeted actor to fine, as there is when the harm is done by a criminal wrongdoer. But if the state should make amends to those it fails to protect from crime, it should equally make amends to those it fails to protect from its own processes when they go awry.¹⁶¹

“[S]tates have no legal obligation to remedy the injuries of the wrongly convicted.”¹⁶² For example, New York’s Unjust Conviction and Imprisonment Act, providing compensation under the state’s Court of Claims Act, “discharge[s] a moral obligation of the State.”¹⁶³ Alberto Lopez suggests that the motivation that underlies crime victim compensation statutes could form a basis for improved state compensation for wrongly convicted persons.¹⁶⁴ If the state feels morally compelled to pass crime victim compensation statutes to “spread out the cost of hardship” on victims harmed by private individuals, it should feel an even stronger moral compulsion “to indemnify the unjustly convicted person and spread out the cost of the harm inflicted” when “it is the failure of the state itself that damaged the victim.”¹⁶⁵

The rights theory that underlies crime victim compensation statutes and carries over to wrongful-conviction compensation statutes “can be based in tort or contract.”¹⁶⁶ Under the contract theory, “citizens have actions for breach of contract if society fails to protect them because they have given up the individual right to exact retribution from a wrongdoer in return for society’s protection.”¹⁶⁷ Under tort theory, “if the State breaches its duty to protect citizens, the injured citizens would have actions against the State for damages in tort.”¹⁶⁸ The rights theory as expressed in wrongful-conviction compensation statutes takes the form of tort, rather than contract, actions.

1. *The Instrumental Theory of Tort Liability*.—The instrumental theory is one of the two main theories underlying tort liability.¹⁶⁹ In the very broadest terms, under the instrumental theory, the threat of liability is a lever to make an actor act in a way that avoids liability. Under this theory, holding a private tortfeasor liable creates in the tortfeasor “an incentive to make cost-justified investments in safety.”¹⁷⁰ Holding a government liable creates “an economic

161. “The reparation of damages caused by erroneous criminal accusations, irrespective of how well founded they seemed, is properly a cost of the operation of the criminal justice system. It is difficult to see why the innocent victims should be forced to absorb this cost.” Boucher, *supra* note 1, at 1101 n.235 (quoting Keith S. Rosenn, *Compensating the Innocent Accused*, 37 OHIO ST. L.J. 705, 716 (1976)).

162. Lopez, *supra* note 4, at 709.

163. Carter v. State, 546 N.Y.S.2d 648, 650 (N.Y. App. Div. 1989).

164. See Lopez, *supra* note 4, at 710-11.

165. *Id.*

166. Mostaghel, *supra* note 156, at 87.

167. *Id.* at 88 (citing Smith, *supra* note 159, at 63).

168. *Id.* (citing Smith, *supra* note 159, at 62).

169. Lawrence Rosenthal, *A Theory of Governmental Damages Liability: Torts, Constitutional Torts, and Takings*, 9 U. PA. J. CONST. L. 797, 798 (2007).

170. *Id.* at 823.

incentive for the government and its officials to make cost-justified investments in preventing constitutional violations.”¹⁷¹ This theory is based on the view that behavior is affected by the existence of economic consequences.

Nevertheless, according to Lawrence Rosenthal, governments are rarely motivated by economic consequences. Rosenthal summarizes the arguments of Professor Daryl Levinson to make this point.¹⁷² Levinson does not believe that the instrumental theory of tort liability applies to governments. He believes that governments respond to political—not financial—incentives, caring more about votes than dollars.¹⁷³ Governments may therefore tolerate behavior that violates the Constitution because that behavior garners political advantage.¹⁷⁴ Levinson gives the example of randomly searching young men in high crime areas—a policy which, despite the likelihood of Fourth Amendment violations, “could also pay such handsome political dividends that liability would have no deterrent effect on elected officials.”¹⁷⁵ Myriad examples spring to mind, including racial profiling to look for drug dealers or ethnic profiling to search airline passengers. Thus, in Levinson’s view, as long as there is a political payoff, governments would rather continue unconstitutional behavior and pay tort damages to plaintiffs injured by those violations.¹⁷⁶

According to Rosenthal, Levinson’s views have generated some negative reactions among academics,¹⁷⁷ but apparently much more approbation.¹⁷⁸ In Rosenthal’s words, “[m]ost academics have been persuaded by Levinson; it has now become fashionable to warn that the consequences of imposing damages liability on government are uncertain at best.”¹⁷⁹ Rosenthal himself disagrees with this conclusion.¹⁸⁰ Based on an examination of the political significance of liability to elected officials, he posits a theory of political behavior that elected officials primarily use the public resources over which they have control in their attempts to gain re-election.¹⁸¹ In attempting to be re-elected, “elected officials

171. *Id.*

172. *Id.* at 824.

173. *Id.* (citing Daryl J. Levinson, *Making Government Pay: Markets, Politics, and the Allocation of Constitutional Costs*, 67 U. CHI. L. REV. 345, 420 (2000)).

174. *Id.* (citing Levinson, *supra* note 173, at 367-68).

175. *Id.* (quoting Levinson, *supra* note 173, at 369-70).

176. *Id.* at 825. Levinson believes that tort liability for constitutional torts could not be imposed on governments under the corrective justice theory either (*see discussion infra* Part III.B.2), since payment would ultimately come from taxpayers. *Id.* He seems to be saying that the moral responsibility would be the government’s, but the government would not deem this responsibility a reason to change behavior because the burden of payment would fall on taxpayers. This is indeed a problem in the funding of current compensation statutes, and I suggest a remedy in Part IV.

177. *Id.* at 827-29.

178. *Id.* at 830.

179. *Id.*

180. *See id.* at 831.

181. *See id.* at 832.

will be highly sensitive to tort liability”¹⁸² because defending against and paying for tort liability takes away resources from other projects.¹⁸³ As a result, elected officials will be “willing to make investments in loss prevention in order to reduce governmental liability costs.”¹⁸⁴

The stumbling block in the path of Rosenthal’s theory could be that governments protect themselves through immunity legislation.¹⁸⁵ Rosenthal addresses this problem. Analyzing a litany of objections to the theory, he ultimately concludes that the ubiquity of immunity legislation¹⁸⁶ shows that government officials are very aware of the liability costs of unconstitutional behavior and want to limit those costs.¹⁸⁷ Thus, he argues, the instrumental justification theory supports tort liability for governments.¹⁸⁸ Even though the desire to minimize costs is primarily accomplished through immunity statutes,

governmental tort liability has an instrumental justification; it creates an incentive on the part of officeholders to allocate resources to loss prevention. There should be a clear political incentive to invest in loss prevention at least when the cost of avoiding an injury is small, the likelihood of injury is great, and the impact on the government’s budget is likely to be large.¹⁸⁹

If there is such an incentive to prevent loss, another question arises: why do states still continue to shield themselves behind immunity laws? Rosenthal’s answer to this question is that it is a matter of costs versus benefits.¹⁹⁰ The political cost of spending money on loss prevention may be greater than the actual “dollars and cents” cost. Thus, the instrumental theory supports tort liability for governments; it is just not a very efficient means of getting governments to take liability for their torts. Indeed, according to Rosenthal, “Professor Levinson was right to claim that governmental tort liability has no efficiency justification comparable to the role of tort liability in the private sector.”¹⁹¹

How does the instrumental theory of tort liability apply to wrongful-conviction compensation statutes? On the one hand, the instrumental theory supports the creation of compensation statutes and a concomitant chipping of the armor of government immunity. Compensation statutes do not seem to carry a high political cost, and they provide several advantages to states: states with compensation statutes can predict costs; they will not be surprised by large

182. *Id.*

183. *Id.*

184. *Id.*

185. See *supra* Part III.A.1 (discussing the Federal Tort Claims Act).

186. See Rosenthal, *supra* note 169, at 838.

187. See *id.* at 841.

188. *Id.* at 842.

189. *Id.*

190. *Id.* at 799.

191. *Id.* at 842.

awards in lawsuits; and there should be fewer wrongful convictions. On the other hand, the instrumental theory of liability does not seem to be a very strong underpinning for current compensation statutes. If it were—that is, if a compensation statute really acted like a lever to make a government actor behave a certain way—then we would see the implementation within these statutes of suggestions designed to change behavior. This is not the case. For example, states have not implemented the suggestion that funding for compensation statutes should come from the prosecutor's budget¹⁹² as a way of curbing overzealous prosecutors. The reason typically given—that the prosecutor in office at the time of a wrongful conviction may no longer be there when the fine is imposed¹⁹³—shows that changing officials' behavior is not a goal of compensation statutes. Because states seem to ignore rather than harness a compensation statute's behavior-changing ability, it appears that most states' compensation statutes are not based on the instrumental theory of tort liability.

2. *The Corrective Justice Theory of Tort Liability.*—Moral rather than economic concerns animate the second major theory of tort liability: that of corrective justice. Under the corrective justice theory, imposing tort damages on individuals “embodies a widely accepted moral obligation on the part of a wrongdoer to make the injured party whole.”¹⁹⁴ Imposing tort damages on the government is “based on an asserted moral entitlement to compensation when one has been the victim of a constitutional wrong.”¹⁹⁵ All of the current wrongful-conviction compensation statutes rest on the corrective justice theory of tort liability.

Corrective justice is the theoretical justification for finding a moral obligation. The vehicle for achieving corrective justice is strict liability—perhaps “the most philosophically and procedurally sound [theory] for imposing

192. See, e.g., Evan J. Mandery, *Efficiency Considerations of Compensating the Wrongfully Convicted*, 41 NO. 3 CRIM. L. BULL., ART. 4, at 1 (2005). Mandery's article opens with a bang: “Here is a simple, seemingly obvious response to the mounting evidence that innocents are being convicted at an intolerable rate: make prosecutors pay for their mistakes regardless of fault.” *Id.*

193. Rosenthal, *supra* note 169, at 835 & 835 nn.152-53.

For example, it is reasonable to believe that the time frame of concern to politicians is the next electoral cycle and that their political judgments are therefore made with only that time frame in mind. For that reason, elected officials might ignore litigation costs or liability exposure, believing that they have no real ability to reduce them quickly enough to affect the current electoral cycle.

Id. (citing James M. Buchanan & Dwight R. Lee, *Tax Rates and Tax Revenues in Political Equilibrium: Some Simple Analytics*, 20 ECON. INQUIRY 344, 345-50 (1982)). On the other hand, Rosenthal continues that “there is reason for skepticism about this view of the time horizons of public officials—most politicians likely plan long careers in public service and will pay a political price if they are still in office when tort judgments must be paid.” *Id.*

194. *Id.* at 823.

195. *Id.* For a discussion of what constitutes constitutional error in wrongful conviction cases, see Garrett, *supra* note 11.

liability on a state.”¹⁹⁶ As explained by Lauren Boucher, “[t]he strict liability theory assumes that in ‘any great undertaking . . . there are bound to be a number of accidents.’”¹⁹⁷ Further,

[b]ecause all citizens benefit from the operation of the criminal justice system (in the form of increased public safety), it is unfair that only one person should bear the cost of an error such as wrongful incarceration simply because he was the unlucky victim of the mistake. Instead, everyone should bear the burden equally.¹⁹⁸

Arguing for the passage of compensation statutes in 1999, Adele Bernhard discussed cases showing “that innocent people have been, and will continue to be, unjustly convicted, as an unfortunate but inevitable consequence of the routine operation of the criminal justice system.”¹⁹⁹ In her view, “neither traditional fault-based tort actions nor civil rights statutes provide a remedy”²⁰⁰ for unjust conviction. Under strict liability, however, there is no need to determine fault. Strict liability works well because

it is the state, through operation of one of its most essential services—the criminal justice system—that has inflicted the harm. Although it may be impossible to hold any individual law enforcement officer, or any particular municipality, liable, the state’s responsibility for the injury is sufficient to generate a moral obligation.²⁰¹

The current wrongful-conviction compensation statutes, all based on strict liability, “do not require claimants to discover why the prosecution was erroneous, or who made mistakes which ‘caused’ the investigation to go awry, or even what those mistakes might have been.”²⁰²

C. Pragmatic Objections to Payment of Compensation Do Not Hold Up

Although theoretical bases support the government’s payment of compensation to the wrongly convicted, those who do not believe that the government must provide such compensation generally present a pragmatic argument to support their view. However, the primary justification offered does not hold up under scrutiny, and strong pragmatic reasons exist to support the payment of compensation. The justification offered for inadequate compensation statutes or for none at all derives from the view that the state cannot right every

196. John J. Johnston, Note, *Reasonover v. Washington: Toward a Just Treatment of the Wrongly Convicted in Missouri*, 68 UMKC L. REV. 411, 414 (2000).

197. Boucher, *supra* note 1, at 1101 (quoting Edwin Borchard, *State Indemnity for Errors of Criminal Justice*, 21 B.U. L. REV. 201, 208 (1941)).

198. *Id.* (citing Borchard, *supra* note 197, at 208).

199. Bernhard, *When Justice Fails*, *supra* note 7, at 92.

200. *Id.*

201. *Id.* at 93.

202. Bernhard, *supra* note 47, at 409.

wrong. “The state cannot be expected to compensate every citizen forced to bear an unjust burden within its boundaries: ‘[c]ertain harms are simply accepted as part of life.’”²⁰³ This is the cost-of-doing-business argument, an argument that accepts that innocent people can and will be imprisoned as an unavoidable aspect of the justice system. The cost-of-doing-business argument should be repudiated for four reasons: (1) it does not save the state money; (2) it leads to treatment that is unfair; (3) it creates incentives to continue the unfair treatment; and (4) it leads states to violate the Constitution.

1. The Cost-of-Doing-Business Argument to Avoid Passing Compensation Statutes Does Not Save the State Money.—Accepting a cost-of-doing-business rationale does not save the state money. The most common justification for the cost-of-doing-business rationale is that it will strain state budgets to provide compensation to wrongfully incarcerated persons who win their freedom.²⁰⁴ Perhaps this fear stems from occasional highly publicized lawsuits that garner million dollar settlements for wrongfully incarcerated individuals.²⁰⁵ It has similarly been feared that compensation schemes for crime victims will create high award and management costs.²⁰⁶ That has not turned out to be the case with crime victim compensation statutes,²⁰⁷ and neither will it be the case with compensation statutes for the wrongfully convicted.

A look at compensation statistics demonstrates that the fiscal argument fails. Gross’s study of exonerations between 1989 and 2003 found that four states—Illinois, New York, Texas, and California—accounted for more than forty percent of the total of 340 exonerations.²⁰⁸ Illinois had fifty-four exonerations, New York had thirty-five, Texas had twenty-eight, and California had twenty-seven.²⁰⁹ Considering just these four states, this is an average of thirty-six exonerations per state over a fourteen year period, which in turn is between two and three per state per year, a number that cannot reasonably be claimed will break the bank.

A look at DNA statistics also bears out that compensation awards will not bankrupt states.²¹⁰ The number of people imprisoned without DNA analysis to

203. Boucher, *supra* note 1, at 1100 n.226 (quoting Bernhard, *When Justice Fails*, *supra* note 7, at 92-93).

204. Fite, *supra* note 8, at 1191 (citing Bernhard, *When Justice Fails*, *supra* note 7, at 105-06).

205. See cases set out in Garrett, *supra* note 11, at 44 n.32 (including some awards of up to \$8.26 million).

206. Bernhard, *When Justice Fails*, *supra* note 7, at 100. Bernhard specifically mentions the opinion of former New York State Assistant District Attorney Richard Kuh, expressed at legislative hearings in New York on crime victim legislation. *Id.*

207. See *id.* at 101. However, this may be because “[r]estitution has been recognized as one of the ‘most underenforced victim rights’ available through the criminal justice system.” Julie Goldscheid, *Crime Victim Compensation in a Post-9/11 World*, 79 TUL. L. REV. 167, 179 (2004).

208. Gross et al., *supra* note 1, at 541.

209. *Id.* at 541 tbl.2.

210. DNA evidence was used to establish innocence in nearly half of the exonerations from 1989 to 2003. Boucher, *supra* note 1, at 1070 (citing Gross et al., *supra* note 1, at 524).

establish their innocence is dwindling—either those cases have now had DNA analysis, or those prisoners are dying. As it becomes more common to do DNA testing on physical evidence at the investigatory stage, according to Bernhard, “the rate of DNA exonerations will inevitably slow. The number of convicted inmates who can locate material, relevant, and untested forensic material will dwindle, as will the number of individuals claiming compensation for unjust conviction.”²¹¹

Along with fearing the cost of implementation, another motivation for states to pass grudging statutes—or none at all—is the fear that a state will compensate people who do not deserve to be compensated.²¹² This fear is also unfounded because “[l]egislatures can carefully draft a statute to prevent unwarranted claims from being considered.”²¹³ Under all compensation statutes, procedural requirements are rigorously enforced.²¹⁴ And “[w]hile statutory provisions are designed to compensate the wrongfully incarcerated, they still require the plaintiff to meet a high burden of proof before compensation will be granted.”²¹⁵ Claimants therefore may not simply restate trial evidence.²¹⁶ The claim of innocence “shall be verified by the claimant;”²¹⁷ this cannot be done with “conclusory and self-serving testimony.”²¹⁸ Rather, claimants must fulfill statutory requirements to establish actual innocence, requirements that function to keep those whose convictions were overturned on procedural grounds from recovering.²¹⁹

A further safeguard for states is the role of trial courts. According to the Iowa Supreme Court, the district court’s “predicate review and assessment of the

211. Bernhard, *Justice Still Fails*, *supra* note 7, at 715.

212. Boucher, *supra* note 1, at 1098.

213. *Id.* at 1099. Boucher also points out that compensation statutes do not result in immediate and unquestioned state compensation, but rather, they create a cause of action. *Id.* She quotes a provision of the Ohio Code to illustrate:

Notwithstanding any provisions of this chapter to the contrary, a wrongfully imprisoned individual has and may file a civil action against the state, in the court of claims, to recover a sum of money as described in this section, because of the individual's wrongful imprisonment. The court of claims shall have exclusive, original jurisdiction over such a civil action.

Id. at 1099 n.221 (quoting OHIO REV. CODE ANN. § 2743.48(D) (West, Westlaw through 2010 legislation)).

214. E.g., *Dvorak v. Pickaway Corr. Inst.*, No. 02AP-452, 2002 WL 31656236, at *3 (Ohio Ct. App. Nov. 26, 2002).

215. Frederick Lawrence, *Declaring Innocence: Use of Declaratory Judgments to Vindicate the Wrongly Convicted*, 18 B.U. PUB. INT. L.J. 391, 394-95 (2009).

216. See, e.g., *Fudger v. State*, 520 N.Y.S.2d 950, 953 (N.Y. App. Div. 1987).

217. E.g., N.Y. CT. CL. ACT § 8-b(4) (McKinney 2003 & Supp. 2011).

218. *Vasquez v. State*, 693 N.Y.S.2d 220, 220 (N.Y. App. Div. 1999).

219. See Michael J. Saks et al., *Toward a Model Act for the Prevention and Remedy of Erroneous Convictions*, 35 NEW ENG. L. REV. 669, 682 (2001); see also discussion *supra* Part III.A.2.a.

claim . . . permits the district court to serve as a gatekeeper.”²²⁰ A claimant may not deprive the court of its gatekeeper role. In an Ohio case, a defendant found to have acted in self-defense was acquitted of a murder charge.²²¹ She argued that under collateral estoppel, she could sue for compensation without establishing actual innocence. The court rejected this argument because acquittal is an essential step that allows a claimant to get to the gate. Only with the acquittal does the court of claims open the gate and give the claimant the opportunity to demonstrate actual innocence.²²²

States whose requirements are very restrictive, as well as states just implementing compensation statutes, should look to states like New York for assurance that less restrictive statutes do not result in floods of claims.²²³ It does not appear that states with well-planned compensation statutes will suffer inordinate strains on their budgets. Similarly, it is unlikely that such states will compensate those who are not entitled to compensation.

2. *The Cost-of-Doing-Business Argument for Refusing to Pass Compensation Statutes Leads to Treatment That Is Unfair.*—It is unfair for the state to treat wrongful conviction and undeserved time in prison as harms that should be “simply accepted as part of life.”²²⁴ It is also unfair to expect people subjected to months or years of undeserved incarceration to return to their lives as if nothing has happened. Moreover, it is wrong for the state to compound the harm of wrongful incarceration by adding another harm—namely, release from incarceration without any provision for a meaningful return to society. Making provision for the wrongfully incarcerated is “[t]he least the community can do.”²²⁵ Indeed, making such provision should be an integral part of criminal justice; “[t]he reparation of damages caused by erroneous criminal accusations . . . is properly a cost of the operation of the criminal justice system. It is difficult to see why the innocent victims should be forced to absorb this cost.”²²⁶ In clarifying the parameters of New York’s compensation statute for the wrongly imprisoned, a New York court said that “the evil sought to be remedied was the likelihood of no recovery.”²²⁷ No state should tolerate the likelihood that a wrongfully convicted person might have no remedy.

3. *The Cost-of-Doing-Business Argument for Wrongful Incarceration Leads to Moral Hazard.*—Accepting the cost-of-doing-business rationale provides no incentive for the justice system to stop practices that lead to unfair treatment. Thus, this rationale leads to moral hazard. Moral hazard is created when there

220. State v. McCoy, 742 N.W.2d 593, 596 (Iowa 2007).

221. Walden v. State, 547 N.E.2d 962, 963 (Ohio 1989).

222. *See id.* at 966-67.

223. *See* Boucher, *supra* note 1, at 1098-99 nn.215-16.

224. *Id.* at 1100 n.226.

225. Bernhard, *When Justice Fails*, *supra* note 7, at 112 (quoting EDWIN M. BORCHARD, *CONVICTING THE INNOCENT: ERRORS OF CRIMINAL JUSTICE* 392 (1932)).

226. Adam I. Kaplan, Comment, *The Case for Comparative Fault in Compensating the Wrongfully Convicted*, 56 UCLAL. REV. 227, 241 n.86 (quoting Rosenn, *supra* note 161, at 716).

227. Carter v. State, 546 N.Y.S.2d 648, 650 (N.Y. App. Div. 1989).

are no consequences for bad behavior. As columnist David Sirota puts it, “without consequences—or worse, with rewards—for wrongdoing, there is an incentive to do wrong.”²²⁸ When the state makes no provision, or only minor provision, for released individuals, the state excuses itself from responsibility for errors and misconduct in criminal investigations, trials, and sentences. No consequences flow from the bad behavior. This lack of consequences effectively tells the criminal justice community that errors and misconduct will be overlooked. Thus, the state’s disinterest creates a moral hazard: an incentive for the criminal justice system to continue doing business as usual. When the state creates the moral hazard of “business-as-usual,” the state will continue to tolerate preventable wrongs. Setting free a wrongfully convicted person is the ultimate expression that preventable wrongs indeed occurred. Under a cost-of-doing-business rationale, there is no incentive to eliminate preventable wrongs.

Courts have a role to play in rejecting the cost-of-doing-business rationale. One way to do so is to ensure that there are consequences for wrongdoing. A number of trial courts have admonished jurors to ignore prosecutors’ improper suggestions to the jury.²²⁹ Appellate courts have reversed and remanded cases because of prosecutors’ misconduct²³⁰ or improper methods.²³¹ Another way courts play such a role is through interpretations of burden-of-proof language in their states’ enabling statutes. Under many statutes, exonerees trying to establish factual innocence must do so by presenting “clear and convincing evidence.”²³² The clear and convincing evidence standard is an elevated standard.²³³ Three statutes, however, direct courts to make allowance for the difficulty of producing clear and convincing evidence when claims are brought years after the original trials.²³⁴ When the statute couples strict requirements with flexibility, courts

228. David Sirota, *Moral Hazards—and Consequences*, S.F. CHRON., Dec. 11, 2009, at A24, available at http://articles.sfgate.com/2009-12-11/opinion/17220728_1_moral-hazard-senate-democrats-bernanke.

229. E.g., *Baker v. State*, 906 A.2d 139, 148-49 (Del. 2006); *Phelps v. State*, 360 N.E.2d 191, 192-93 (Ind. 1977). In *Helleston v. State*, 5 S.W.3d 393, 397 (Tex. Crim. App. 1999), the prosecutor was told he could mention parole law but could not discuss it in regard to the sentence to be determined. When he did, the court told the jury to disregard the prosecutor’s remarks. Courts do this routinely, but if not carefully worded, the court’s warning to the jury might actually make the problem worse by highlighting the improper remarks. *People v. Bolden*, 589 P.2d 396, 400 n.5 (Cal. 1979).

230. E.g., *DeFreitas v. State*, 701 So. 2d 593 (Fla. Dist. Ct. App. 1997).

231. *Schoels v. State*, 966 P.2d 735, 743 (Nev. 1998) (Springer, C.J., dissenting).

232. E.g., ME. REV. STAT., tit. 14, § 8241(2) (West, Westlaw through 2009 2d Reg. Sess.); N.J. STAT. ANN. § 52:4C-3 (2011); W. VA. CODE ANN. § 14-2-13a(f)(1) (West, Westlaw through 2011 Reg. Sess.).

233. “The clear and convincing evidence standard is somewhere between the preponderance standard of civil cases and the reasonable doubt standard of criminal cases.” *In re G.B.R.*, 953 S.W.2d 391, 396 (Tex Crim. App. 1997).

234. N.Y. CT. CL. ACT § 8-b(1) (McKinney 2003 & Supp. 2011); N.J. STAT. ANN. § 52:4C-1; W. VA. CODE ANN. § 14-2-13a(f).

have the tools to issue decisions that whittle away at the cost-of-doing-business rationale.²³⁵ When statutes are not explicit, a court's understanding of legislative intent may be similarly important. For example, a Louisiana court's interpretation was influenced by its understanding that legislative intent did not limit "the introduction of evidence related *in any way* to the conviction and the proof of factual innocence."²³⁶ When statutes are not explicit, courts may facilitate consequences for wrongdoing by adopting a preponderance of the evidence standard instead of a clear and convincing evidence standard.²³⁷

4. The Cost-of-Doing Business Argument for Wrongful Incarceration Leads States to Violate the Constitution.—Finally, accepting the argument that wrongful incarceration is a cost of doing business leads to the ultimate moral hazard: the continued acceptance of constitutional violations. In a criminal appeal based on constitutional error, the court's focus is "on whether evidence of guilt could excuse constitutional error."²³⁸ In effect, such a focus means that the court is balancing the possibility that the accused is guilty against the certainty that a constitutional error was committed. Rather than aiming for the elimination of constitutional errors from criminal trials, this focus invites the routine repetition of constitutional errors.

Whenever an exoneree succeeds in winning release after years of wrongful confinement, sympathy grows among the public for the payment of compensation. Tort-based compensation statutes not only respond to this sympathy, but also may lead to criminal justice reforms in such areas as mistaken

235. The New Jersey statute, in discussing the burden of proof that exonerees bear in establishing factual innocence, states that

it is the intent of the Legislature that the court, in exercising its discretion as permitted by law regarding the weight and admissibility of evidence submitted pursuant to this section, may, in the interest of justice, give due consideration to difficulties of proof caused by the passage of time, the death or unavailability of witnesses, the destruction of evidence, or other factors not caused by such persons or those acting on their behalf.

N.J. STAT. ANN. § 52:4C-1.

236. *In re Williams*, 984 So. 2d 789, 793 (La. Ct. App. 2008) (emphasis added). Although the result—vacatur of the trial court's award of compensation—was bad for the exoneree, the case clarifies the burden of proof and the steps in establishing proof of innocence.

237. *Walden v. State*, 547 N.E.2d 962 (Ohio 1989) (applying OHIO REV. CODE ANN. § 2305.02 (West, Westlaw through 2010 legislation)); *Duncan v. State*, No. 5625, 1972 WL 16790 (Ill. Ct. Cl. Apr. 12, 1972) (applying 37 ILL. COMP. STAT. ANN. § 439.8(c) (1967) and interpreting how to prove innocence of the fact of a crime). A claimant seeking to prove an unjust conviction claim carries a heavy burden of "evidence that is neither equivocal nor open to opposing presumptions." *Solomon v. State*, 541 N.Y.S.2d 384, 385 (N.Y. App. Div. 1989).

238. Garrett, *supra* note 11, at 38. In the introduction to this article, Garrett explains how the development of harmless error rules in criminal cases makes it hard for an accused to establish innocence. *Id.* He says that wrongful conviction lawsuits based on tort liability may have a dramatic reformative effect because they may change the "truth-defeating nature" of aspects of criminal procedure. *Id.* at 36.

eyewitness identification,²³⁹ crime lab error,²⁴⁰ and ineffective assistance of counsel.²⁴¹ Reforms in these areas cut down on the potential for constitutional violations. When constitutional violations are perceived to carry a cost, justification for the cost of doing business evaporates.

A wrongful-conviction compensation statute creates incentives that make it worthwhile for state actors to conform their conduct to the Constitution. Assuming that a compensation statute makes a department of government responsible for its constitutional violations, the motivation of that department to overlook such violations may diminish. Unconstitutional practices that a department might previously have ignored may no longer seem justifiable when violations leading to wrongful convictions have consequences for those who committed them.

An argument can be made that imposing consequences on government actors for unconstitutional behavior will tie the government's hands. However, consequences need not be unlimited. Rosenthal explains this in political terms. According to Rosenthal, a constitutional right should not depend on whether the right is politically acceptable:

Inherent in the concept of a constitutional right is that its protection does not depend on the political acceptance of the right at stake. Thus, political accountability is an unacceptable method for securing constitutional rights; the Constitution protects even the unpopular or politically inexpedient. . . . [A] law of constitutional torts must place pressure on the government to conform all of its conduct to the Constitution. That does not imply, however, that damages are always properly awarded for a constitutional violation. Once one understands that the primary virtue of damages awards against the government is to create a political incentive to undertake loss prevention, there is ample room for damages-limiting doctrines that protect the interests of the taxpayers and avoid unwarranted reallocation of scarce public resources.²⁴²

This means that passing wrongful-conviction compensation statutes in no way ties a government's hands.

Passing wrongful-conviction compensation statutes will create political incentives for loss prevention, i.e., improving criminal justice practices, if part of the funding for the damage awards comes from the departments or offices that have the most ability to prevent constitutional violations. Some part of the damage award must come from police and prosecutor budgets because their actions directly create or prevent these violations. But statutory damage caps are acceptable because they "preserve political pressure on government to conform

239. One among many articles dealing with criminal justice reform is Boucher, *supra* note 1, at 1074-78.

240. *Id.* at 1078-79.

241. *Id.* at 1080-82.

242. Rosenthal, *supra* note 169, at 856 (internal citation omitted).

its conduct to the law but mitigate the anomalies associated with governmental damages liability.”²⁴³ The aim of damage statutes should be “[a] regime of limited liability that nevertheless imposes a sufficient political price to minimize the likelihood of constitutional violations.”²⁴⁴ Translated to the compensation issue, this means that the government that created the harm then passes a statute to give damages for the harm. Because the government must continue to fulfill its other roles, the remedy will be limited. But a primary goal of the statute is to minimize the likelihood that the government will convict and imprison innocent persons. Therefore, the remedy should inflict some pain that matters on the “bad” actors. Suggestions such as fining or charging the prosecutor who brought the case will not work because that prosecutor may be gone, and no other prosecutor will see the remedy as cautionary.²⁴⁵ But taking some portion of the compensation money from the budget of prosecutors would exact “a sufficient political price”²⁴⁶ because any prosecutor could expect to lose funding whenever damages are awarded.

D. Pragmatic Reasons Support the Payment of Compensation

Refusing to pass generous wrongful-conviction compensation statutes has no practical justification. Looked at pragmatically, a refusal to pay compensation based on the cost-of-doing-business rationale cannot be justified. Paying compensation does not strain state budgets—and not only is refusing to pay compensation unfair, it also creates disincentives to improve bad practices and leads to constitutional violations. Further, pragmatic reasons exist for putting fair and generous compensation statutes into place.

First, passage of statutes contributes to a rebalancing of priorities. If the state does not take adequate steps to compensate the individual harmed by its processes, it demonstrates that it lacks responsibility for the errors. The state’s demonstration that it lacks responsibility allows and even encourages further

243. *Id.* at 863. Rosenthal does not illustrate anomalies. A possible example is seen in *Rooney v. United States*, 634 F.2d 1238 (9th Cir. 1980). Plaintiff Rooney was injured working for a contractor hired by the federal government. All three parties were negligent, but the government argued it could not be liable for damages “apportionable to its joint tortfeasor, the contractor.” *Id.* at 1245. The court said that “[u]nder California law, each concurrent tortfeasor in a . . . comparative negligence case is jointly liable with the others. This rule applies regardless of the apportionability of negligence . . . Since each defendant’s negligence was a proximate cause of Rooney’s indivisible injuries, each defendant is liable for the full amount of damages.” *Id.* (internal citations omitted). Thus, the court found the government responsible for all damages except the thirty percent attributable to the comparative negligence of the plaintiff. The result, according to the court, is that “[w]e are faced with an anomaly which we must accept: the [g]overnment, which is the least culpable of the three of the negligent parties, will bear the greatest burden in damages.” *Id.* The court affirmed judgment for the plaintiff.

244. Rosenthal, *supra* note 169, at 863.

245. See generally Medwed, *supra* note 9.

246. Rosenthal, *supra* note 169, at 863.

wrongdoing. Putting wrongful conviction statutes into place signals that the state is aware it may make mistakes and that those mistakes may be grave. It also signals that those who either commit or do not prevent mistakes in criminal process shall bear some responsibility for wrongful convictions. Making it easier for an exoneree to receive compensation through statute should result in fewer false convictions since prosecutors' offices may be motivated to pay more attention to fair trial techniques.²⁴⁷ In fact, compensation statutes should be designed to make it easier, not harder, for exonerees to receive compensation.²⁴⁸ Draconian compensation statutes rest on the argument that a restrictive statute will lead to few awards because it is so hard for exonerees to qualify. That rationale does not comport with our fundamental understanding of a justice system that protects the innocent from the ultimate deprivations: those of life and liberty.

Second, compensation statutes allow governments to forecast costs. Errors will occur, which explains the cost-of-doing-business argument. But tolerating errors in the criminal justice system under a cost-of-doing-business rationale may turn out to be costly for the state as well as unfair to the exoneree whose harm is justified under this rationale. Implementing compensation statutes allows the state to estimate its costs and not leave itself open to unexpected and potentially high damage awards. Passing compensation statutes that streamline the process of applying for compensation may lead to lower costs since the state will presumably be defending against fewer exoneree common law tort suits against prosecutors and § 1983 suits against municipalities and the police. Passing compensation statutes that make it easier for exonerees to receive compensation will enable the state to set up a predictable cost structure. Refusing to compensate exonerees as a cost-containing measure may backfire since sympathetic juries are likely to give large damage awards to the wrongfully convicted. For instance, the City of Chicago paid \$1.5 million to a man wrongly convicted of murder after a city crime lab analyst was found to have falsified reports.²⁴⁹ In another Chicago case, the city agreed to pay \$9 million to settle a wrongful rape conviction lawsuit.²⁵⁰ Similarly, New York City paid \$5 million when it came to light that the prosecution had suppressed exculpatory evidence in a rape case.²⁵¹

247. See Garrett, *supra* note 11, at 53-56, 71-99.

248. For exonerees, the process should be more like submitting an application than filing a lawsuit. Making the process an administrative rather than adversarial proceeding will reduce the toll on exonerees and the costs to government.

249. See Steve Mills et al., *When Labs Falter, Defendants Pay*, CHI. TRIB., Oct. 20, 2004, available at <http://truthinjustice.org/labs-falter.htm>.

250. Maurice Possley & Gary Washburn, *City Will Pay \$9 Million in False Jailing*, CHI. TRIB., Jan. 28, 2006, at 1, available at http://articles.chicagotribune.com/2006-01-28/news/0601280288_1_dna-testing-chicago-police-supt-false-confessions.

251. Andrea Elliott, *City Gives \$5 Million to Man Wrongly Imprisoned in Child's Rape*, N.Y. TIMES, Dec. 16, 2003, at B3, available at <http://nytimes.com/2003/12/16/nyregion/city-gives-5-million-to-man-wrongfully-imprisoned-in-child-s-rape.html>.

If, instead of reacting to lawsuits, states develop plans for paying compensation under wrongful-conviction compensation statutes, they can encourage procedural changes that will lead to fewer instances of wrongful arrest and conviction. At the same time, they can prepare for the predictable costs of the scheme.

IV. FUNDING COMPENSATION STATUTES

A state may refuse to pass a compensation statute for the wrongfully convicted. The state does not, however, escape paying for the harm it has tolerated, if not actually caused. As difficult as it is for exonerees to sue on wrongful imprisonment grounds, exonerees do win lawsuits, and when they do, their damage awards are often significant.²⁵² Some entity of government pays when an exoneree wins a judgment against a prosecutor or a municipality. States should therefore accept their responsibility for wrongful convictions and implement fair and fairly administered compensation schemes. Recognizing that exonerations will occur means that the state can prepare for the inevitable payment of compensation instead of trying to find ways to deny recovery to wrongfully imprisoned persons.

Rosenthal points out that many proposals calling for governments to pay tort damages require the government

to assume costs in order to avoid losses experienced by others; consequently, they all have a negative impact on government budgets, regardless of the externalized benefits they may produce—unless one can make the rather implausible claim that these proposals would be so popular that the voters would tolerate an increase in taxes to fund the new expenditures that they necessitate. Yet one cannot find in any of the proposals for new governmental liability any consideration of the consequences that new liabilities will have on government budgeting, or on those who depend on government budgets for the variety of social goods allocated through that process.²⁵³

By consequences, Rosenthal means that increased government spending in one area will result in less money for government spending in other areas. It is not enough for states to pass compensation statutes; they must also determine how these statutes will be funded.

Current wrongful-conviction compensation statutes do not pay much attention to the sources of funding. In California, for example, exonerees are not discussed explicitly in budget information on the state's website.²⁵⁴ California

252. See Robert T. Garrett, *Texas House Votes to Boost Compensation for Wrongly Imprisoned*, DALL. MORNING NEWS, Apr. 25, 2009, available at http://www.dallasnews.com/sharedcontent/dws/dn/dnacases/stories/DN-innocent_25tex.ART.State.Edition1.4a9c71a.html.

253. Rosenthal, *supra* note 169, at 845.

254. See generally Letter from Dep't of Finance to Dep't of Budget Officers, Dep't Accounting Officers, & Dep't of Finance Budget Staff (Sept. 28, 2010), available at

funds its wrongful-conviction compensation statute from the state's general fund²⁵⁵ pursuant to the California Penal Code.²⁵⁶ A specific source within the general fund is not designated. The state's department of corrections receives its funding from the general fund.²⁵⁷ Perhaps funding comes from the corrections budget, perhaps not. The general fund is a fairly haphazard source, as the provision of funds for DNA testing illustrates. In 2001,

California mandated that all costs associated with representing inmates pursuant to Penal Code section 1405 to investigate and, if appropriate, file motions for DNA testing of biological evidence where such testing could prove innocence, *be borne by the State*. In that same year, California allocated \$1.6 million dollars over two years to provide counsel to assist inmates with innocence claims. For 2002 and 2003, the NCIP [Northern California Innocence Project] and CIP [California Innocence Project] received state funding. That funding *was discontinued as a result of state budget cuts in 2003.*²⁵⁸

Other states have equally vague sources of funding. Alabama pays awards out of any available state funds.²⁵⁹ New Hampshire pays claims from any money in the treasury not otherwise appropriated.²⁶⁰ North Carolina provides awards from its contingency and emergency fund or other available state funds.²⁶¹ These states' statutes at least mention some funding source. Most, however, are silent. The most completely thought out statute seems to be Maryland's, which specifies that the board of public works should make payments to exonerees using money in the general emergency fund or money provided by the governor in the annual budget.²⁶² If governments must "assume costs" in an attempt to avoid losses that others experience at their hands, they need to pass compensation statutes whose funding sources will be secure.

Assuming a state is revisiting its statute or creating a new one, how can it fund the statute? One way to pay for harm is through insurance. In an insurance scheme, the insured protects against a risk by paying premiums to an insurer to

http://www.dof.ca.gov/budgeting/budget_letters/documents/BL10-26.pdf.

255. CAL. PENAL CODE §§ 4900-06 (2010).

256. *Id.* § 2713.1.

257. KAISER FAMILY FOUND., CALIFORNIA: DISTRIBUTION OF STATE GENERAL FUND EXPENDITURES (IN MILLIONS), SFY2008, <http://www.statehealthfacts.org/profileind.jsp?rgn=6&ind=33&cat=1> (last visited Feb. 16, 2011). In fiscal year 2008, the department of corrections received 9.4% of the general fund. *Id.*

258. CAL. COMM'N ON THE FAIR ADMIN. OF JUSTICE, FINAL REPORT 107 (Feb. 22, 2008), available at <http://www.ccfaj.org/documents/CCFAJFinalReport.pdf> (emphases added).

259. ALA. CODE § 29-2-160(a) (2003).

260. N.H. REV. STAT. § 541-B:13 (2007).

261. N.C. GEN. STAT. § 148-84 (LEXIS through 2010 Reg. Sess.).

262. MD. CODE ANN. STATE FIN. & PROC. § 10-501(a)(2) (West, Westlaw through 2010 Reg. Sess.).

cover that risk.²⁶³ Employers provide insurance “to account for the risk of liability that employees face.”²⁶⁴ Governments offer insurance to public employees through statutes, policies, or collective bargaining agreements.²⁶⁵ Consumers buy insurance to extend product warranties. They buy trip insurance to minimize losses from having to cancel a trip. Self-employed workers buy medical insurance. Fathers and mothers buy life insurance to safeguard their children if the parents die.²⁶⁶ In all of these contexts, the purchaser of the insurance buys the insurance because he understands that a danger of harm exists generally and that the harm may befall him specifically.²⁶⁷ Yet even with sensational news stories about people being set free after years and years of wrongful imprisonment, there is no wrongful conviction insurance. Discussing ways in which requiring compensation for the wrongly convicted should lead to improvement in the criminal justice system, Evan Mandery says that

[i]nsurance theory suggests that in the open market consumers would not choose to purchase “constitutional tort insurance.” Generally speaking, people choose to buy insurance against losses that reduce wealth, but do not insure against intangible harms, such as emotional distress or affronts to . . . [dignity], that have no direct or indirect effect on wealth. Since most constitutional torts cause intangible damages, insurance theory argues that requiring citizens to purchase insurance against these kinds of injuries reduces their net welfare.²⁶⁸

Arguably, wrongful incarceration is a constitutional tort that reduces net wealth and inflicts an array of intangible damages. But the point is that people will most likely not voluntarily spend money, and thereby reduce their net worth, to insure themselves against an event that seems so remote from the average person’s experience.

Assuming individuals will not buy wrongful conviction insurance, but realizing that individuals are harmed by wrongful conviction (generally through no fault of their own), should states impose higher taxes on citizens to spread out the risk? In the case of strict products liability,

consumers pay premiums to manufacturers—through higher product prices—in exchange for indemnification against injuries caused by the products. The thought is that manufacturers are better able to bear the costs of injuries than individual consumers. . . . [T]he premium . . . must be one that they [consumers] would choose to pay in a private market.²⁶⁹

263. See Rosenthal, *supra* note 169, at 820-21.

264. *Id.* at 820.

265. *Id.* at 819.

266. Mandery, *supra* note 192, at 5.

267. I do not include automobile insurance in this discussion because its utility to the public is readily apparent, unlike insurance for something like wrongful convictions.

268. Mandery, *supra* note 192, at 7.

269. *Id.*

The analogy is between consumers paying premiums to manufacturers through higher prices and citizens paying premiums to the government through higher taxes. Governments are unlikely to see raising taxes as a solution because raising taxes imposes a political cost on elected government officials. Voters may not re-elect officials who raise taxes. A government will therefore avoid raising taxes if those taxes are to insure against a constitutional harm that is typically not the kind of harm people will insure against with their own money.

If states are not willing to pay for compensation statutes through direct taxation, is there some other way for states to assume this burden? A more useful analogy is to see states as manufacturers insuring themselves against injuries caused by their product, the criminal justice system. Manufacturers are thought to be better able to “to bear the costs of injuries than individual consumers.”²⁷⁰ But manufacturers can bear these costs because they get them back from consumers in the form of higher prices. Since the states do not want to raise taxes to pay for wrongful convictions, they need an insurance company of their own. So the analogy shifts again: now each individual state is a consumer, and the group of all the states becomes the insurer. Banding together, the states could create a wrongful convictions funding pool much larger than any single state’s fund for damage awards. All the states would contribute,²⁷¹ but only a few states would likely need to use the funds in any particular year. In this manner, the funds will be able to grow.

A state could also create its own fund to cover the costs of compensating exonerees. Ideally, the fund would grow large enough that compensation could be paid from the interest it generates, and not from the principal. An argument can be made that there is no need to create a dedicated fund to cover the costs of exoneree compensation because specific awards “are a small percentage of any state’s annual budget.”²⁷² Looking just at numbers, this seems to be a valid point. For example, California’s total budget expenditure for 2008 was about one hundred billion dollars.²⁷³ An award of one or two million is nothing in the face of that number. Would this amount be more significant if it is targeted to come out of the corrections budget instead of the general fund? Even a small award makes up a larger percent of the corrections budget than the overall state budget. California’s corrections budget for 2008 was 9.4% of the state’s total budget.²⁷⁴ This figure is certainly a small percentage of the whole state budget; however, it still amounted to about \$9.6 billion. When corrections spends over \$9 billion

270. *Id.*

271. States could contribute based on the size of their felony prison populations, for example.

272. Karin D. Martin, *A Model State Policy for the Treatment of the Wrongfully Convicted*, LIFE AFTER EXONERATION PROGRAM 31 (2006), <http://www.exonerated.org/content/images/articles/model%20state%20policy%20-%20karin%20martin.pdf>. Martin conducted this study as part of the requirements for the Master of Public Policy degree of the Goldman School of Public Policy, Univ. of Cal., Berkeley. *Id.* at 1.

273. KAISER FAMILY FOUND., *supra* note 257.

274. *Id.*

a year, an award of \$1 million or \$2 million should be affordable. But this argument—that there is no need to create a dedicated fund because specific awards are relatively small—must be a spurious argument because some states do not provide compensation at all, and those that do have compensation statutes do not spontaneously make awards. Instead, they force the exoneree to sue the state to set the compensation machinery in motion. Therefore, the insignificance of the amount does not guarantee payment.

Assuming that a state decides to create a fund to pay for compensation claims, the question is how the state will make payments into the fund. Three sources can be tapped. The first source of funds is prosecutors' budgets. Whenever an exoneree is awarded compensation for wrongful incarceration, the attorney general's office should be required to contribute to the fund.²⁷⁵ Even if the process for establishing factual innocence reveals that the state was without fault, the attorney general's office must still make the payment to the fund.²⁷⁶ Also, in those few states that have implemented administrative procedures instead of adversarial ones,²⁷⁷ the attorney general's office should still be required to pay into the fund.

The second source of funds is the corrections budget. The fund could be set up to require a yearly payment from the corrections budget, which could be a set amount or a fluctuating percentage. In either case, it can be calculated in various ways. The amount could be based on the size of the prison population. Alternatively, it could be based on the number of felons in prison or the average number of exonerations per year over a set period. In California, for example, twenty-seven claims for exoneration were approved between 1989 and 2003.²⁷⁸ Taking the ratio to be twenty-seven exonerations per fifteen years, the average

275. This proposal may raise the ire of prosecutors. *See* Medwed, *supra* note 9, at 157.

276. An example is “if the erroneous conviction were entirely the fault of a perjurious witness.” Saks et al., *supra* note 219, at 682 n.35. Incidentally, the state may bring criminal charges against the perjurious witness. *Id.*

277. Alabama’s remedy is exclusively administrative. *See* ALA. CODE § 29-2-151 to -165 (2003). California has an administrative scheme. *See* CAL. PENAL CODE § 4900 (2010). North Carolina and Wisconsin also have administrative schemes. *See* N.C. GEN. STAT. § 148-83 (LEXIS through 2010 Leg. Sess.); WIS. STAT. ANN. § 775.05 (West, Westlaw through 2009 Reg. Leg.). In Texas, an exoneree applies for administrative remedies. *See* TEX. CIV. PRAC. & REM. CODE ANN. § 103.051 (West, Westlaw through 2009 Reg. Leg.). In New Hampshire, for a claim for less than \$5000, the claimant follows administrative procedures. For a claim between \$5000 and \$50,000, there is concurrent jurisdiction between the administrative and judicial fora. Above \$50,000, the claimant must follow the judicial route and sue in superior court. N.H. REV. STAT. ANN. § 541-B:9 (2007). In New Hampshire, then, most wrongful compensation claims would be brought as suits against the state in superior court. In Tennessee, the board of claims investigates and hears administrative claims for compensation. TENN. CODE ANN. § 9-8-108(a)(7) (West, Westlaw through 2010 Reg. Sess.).

278. CAL. COMM’N ON THE FAIR ADMIN. OF JUSTICE, *supra* note 258. In addition, twenty-five claims were denied and nineteen were dismissed as untimely, incomplete, or because the claimant was still in prison. *Id.*

yearly number of exonerations would be 1.8, or about two exonerations per year.

The third source of revenue comes from re-visioning how the department of corrections spends money to maintain inmates. The state's correctional authority must be seen as establishing a "line" in the budget for each prisoner. This line should not end with a prisoner's exoneration. Each year, the money that would have been spent on the (now-exonerated) prisoner would go into the exoneree's compensation fund. The exonerated prisoner would receive the set statutory amount annually until he has received payment for each of the years that he served in prison. If an exoneree dies before receiving his full payout, the remainder of the statutory amount would go to his heirs annually until the exoneree's years of prison service are compensated. Even after the exoneree or his estate has received the full payout to which he is entitled, however, the exoneree's budget line need not evaporate. In order to build up the fund, corrections could continue to make payments from the exoneree's prison budget line until the exoneree actually dies or until he was projected to die, whichever is longer. For a life sentence, payment should continue for the projected actuarial life span of the exonerated person.

All three of the proposed revenue sources for a state's own compensation fund are based on the two theories underlying tort liability: (1) the corrective justice theory that the state has a moral obligation to make victims whole²⁷⁹ and (2) the instrumental theory that the threat of liability creates incentives to avoid behavior that triggers liability.²⁸⁰ In corrective justice terms, it is appropriate to impose some of the moral obligation to make victims whole on prosecutors' offices since those offices bear some responsibility for the harm. Making the office strictly liable under the instrumental theory obviates the necessity to find any particular culprit, which is useful if specific prosecutors are no longer in office.²⁸¹ Even though the prosecutor who allowed a wrongful conviction to occur may have left the department, every prosecutor taking office after passage of this statute will know that a portion of the funding for exoneree awards comes from the prosecutor's budget.

As noted earlier in this Article, prosecutors will resist this requirement.²⁸² Describing existing statutes, Medwed writes that "[a]lthough these statutes do not expressly designate that funds used for this compensation should be drawn directly from prosecutors' budgets, the impact of these payouts on state coffers could conceivably have an indirect effect on the amount of money allocated to prosecutors partially dependent on state funding."²⁸³ I suggest that the funding under revised statutes should have a direct effect on money allocated to

279. See *supra* Part III.B.2.

280. See *supra* Part III.B.1.

281. Cf. Medwed, *supra* note 9, at 144 (assigning post-conviction motions based on whether lawyer who prosecuted case was still in office).

282. See *supra* note 275. I realize that this proposal implicates the very reason that prosecutors may resist post-conviction claims of innocence. However, all participants in wrongful convictions should participate in the remedy.

283. Medwed, *supra* note 9, at 157 (internal footnotes omitted).

prosecutors. Knowing that their budgets will be affected every time a wrongfully convicted person is exonerated should create a stronger incentive for prosecutors to avoid the practices that give rise to wrongful convictions.²⁸⁴

The second source of revenue—the automatic yearly payment from the corrections budget to the compensation fund—works similarly. It is not possible at the level of a huge department to blame specific individuals for the miscarriages of justice that result in wrongful convictions. But holding that department strictly liable for a significant portion of compensation funding is appropriate because that is the only department that could have condoned bad practices and can now stop them.

The third source of revenue—requiring corrections to continue carrying the expense of maintaining a prisoner by paying into the fund what it would have spent if the prisoner had not been exonerated—similarly satisfies both theories of tort liability. Continuing to “pay” for a prisoner who has left the system helps provide a viable fund that could make future victims whole. Requiring corrections to make these payments puts pressure on the department of government that is both most likely and most able to make changes, leading to fewer wrongful convictions.

There is a difference between the first two revenue streams and the department of corrections’ budget “line.” The first two revenue streams will yield reasonably predictable amounts because the amounts are based on averages. The method of calculation (and re-calculation, from time to time) of these averages must be established in the statute. By contrast, the third source, which is the amount the state pays to the fund under the budget line concept, is more fluid. That amount depends on several variables. One variable is how many exonerated individuals are receiving compensation at any one time. The more exonerees receiving compensation, the higher the amount coming from budget lines will be. The second variable is whether an exoneree was on death row when he was exonerated and released. The cost to the state to maintain a person on death row is higher than the cost to maintain a prisoner serving a life sentence.²⁸⁵ When a death row inmate is exonerated, therefore, the state would put into the fund the difference between what it costs to maintain a death row inmate and an inmate serving a life sentence without the possibility of parole.²⁸⁶

284. The creation of incentives to avoid bad practices is a third pragmatic reason in favor of wrongful-conviction compensation statutes. *See supra* Part III.D.

285. A New Jersey commission studying the death penalty found that the state would save \$1.3 million per year in costs of incarceration if it switched just one death row inmate’s sentence to a sentence of life without parole. Joe Bargmann, *Debating the Cost of Capital Punishment*, PARADE, Jan. 31, 2010, at 6, available at <http://parade.com/news/intelligence-report/archive/100131-debating-the-cost-of-capital-punishment.html>. Natasha Minsker, death penalty policy director of Northern California’s ACLU affiliate, has said that changing death sentences to permanent imprisonment would save California \$1 billion over five years. Scott Smith, *State Moves Closer to Resuming Executions*, STOCKTON REC., Jan. 6, 2010, available at http://www.recordnet.com/apps/pbcs.dll/article?AID=/20100106/A_NEWS/1060309.

286. In a generous statute, the compensation for an exonerated death row inmate would be

But even allowing for a higher level of compensation because of the traumatic experience of facing execution, the annual cost of compensation for an exonerated death row inmate could be lower than the annual cost to maintain that person on death row.

Unpredictable results may also arise in the case of older prisoners. Thus, the age of the prisoner is the third variable. The cost to the state of paying into the fund for elderly exonerees is probably going to be less than the cost of maintaining them as older prisoners. Although it is true that "like other segments of the population, inmates are living longer,"²⁸⁷ many criminology researchers consider offenders to be "elderly" at fifty-five.²⁸⁸ According to Snyder, this is because of "the shorter life expectancy and lower health status of criminal offenders."²⁸⁹ If inmates are elderly at fifty-five, with impaired health and shortened life expectancy, they can be expected to develop the illnesses of age sooner than those who are not inmates. Older prisoners need more medical care; as a result, they cost more to maintain.²⁹⁰ If a prisoner is exonerated before developing the illnesses of age, it is possible that the budget line payment that the state continues to make into the fund every year on his behalf will be less than the actual cost to maintain the prisoner.

CONCLUSION

Wrongful convictions cost states money, resources, public goodwill, and moral authority. Well-planned wrongful-conviction compensation statutes help to mitigate all of these costs. But whether states contribute to a multi-state pool or create their own dedicated funds, state costs will certainly remain. Initially, states will need to spend money to get their funds operational. And states will have to commit continuing resources to maintaining the funds. Nevertheless, passing wrongful-conviction compensation statutes that include identified funding sources has important advantages over the haphazard systems generally in place now. The human cost to exonerees will be less; the financial cost to the states will be more predictable and more manageable.

Of course, it is better for exonerees not to have to fight yet again to receive compensation. It is better for states, too. Their criminal processes will improve. Their reputations will not be savaged in the press when juries make huge awards to exonerees. Presumably, it will be easier for states to fulfill their moral obligations to their citizens. And a citizenry often cynical about the justice system may begin to find renewed optimism that there actually is justice in the system.

adjusted to reflect that the exoneree was on death row.

287. Cindy Snyder et al., *Older Adult Inmates: The Challenge for Social Work*, 54 SOC. WORK 117, 117 (2009).

288. *Id.*

289. *Id.*

290. See Jonathan Turley, *Older Prisoners and Overcrowding*, RES IPSA LOQUITUR BLOG (Dec. 6, 2007), <http://jonathanturley.org/2007/12/06/older-prisoners-and-overcrowding/>.

Indiana Law Review

Volume 44

2011

Number 2

NOTES

WHO NEEDS CONTRACT LAW?—A CRITICAL LOOK AT CONTRACTUAL INDEMNIFICATION (OR LACK THEREOF) IN FHAA AND ADA “DESIGN AND CONSTRUCT” CASES

CHARLES DAUGHERTY*

INTRODUCTION

The right to contract freely with the expectation that the contract shall endure according to its terms is as fundamental to our society as the right to write and to speak without restraint. Responsibility for the exercise, however improvident, of that right is one of the roots of its preservation.

A rule of law which would sanction the renunciation of a bargain purchased in freedom from illegal purpose, deception, duress, or even from misapprehension or unequal advantage leads inexorably to individual irresponsibility, social instability and multifarious litigation.¹

For generations in the United States there has been a debate surrounding the importance of freedom of contract.² Many scholars link freedom of contract with notions of individualism, democracy, and free will.³ Others note that enforcement of obligations freely bargained for is essential for a capitalist economy, and linking freedom of contract to treasured American values preserves the hierarchical structure of a capitalist system.⁴ Within this debate, governmental

* J.D. Candidate, 2011, Indiana University School of Law—Indianapolis; B.A., 2007, Purdue University, West Lafayette, Indiana. Recipient of the Papke Prize for Best Note in Volume 44, endowed by and named in honor of David R. Papke, former R. Bruce Townsend Professor of Law and faculty advisor to the *Indiana Law Review*.

I would like to thank Professors Antony Page and Cynthia Adams for providing invaluable insight and guidance throughout the writing process. I would also like to thank Danielle Tucker and Katie Boren for their constructive feedback and helpful suggestions. Lastly, I would like to thank my family as well as Billy Holland and Caleb Singleton for their undying support and inspiration.

1. Blount v. Smith, 231 N.E.2d 301, 305-06 (Ohio 1967) (citation omitted).

2. See Samuel Williston, *Freedom of Contract*, 6 CORNELL L.Q. 365, 365-67 (1921) (discussing the changing perception of freedom of contract over time).

3. See *id.* at 366.

4. See generally KARL MARX, THE GERMAN IDEOLOGY (1932), reprinted in KARL MARX:

intervention is the key variable that splits opinions.⁵ Milton Friedman and John Kenneth Galbraith, two highly influential economists,⁶ nicely represent the diametric views of free market versus governmental intervention.⁷ Friedman advocated freedom for market participants to allocate resources and responsibilities with little governmental intervention in order to promote the most efficient economic system.⁸ Galbraith advocated more governmental intervention in the market to help decide through regulation what is most efficient and productive.⁹ Both sides of the debate recognize to some degree that the ability of parties, particularly business entities, to freely negotiate a deal with the expectation that the deal's terms will be upheld by courts is essential for carrying on business.¹⁰ Parties often come to the negotiating table with a similar profit-making goal but different plans of how best to achieve that goal.¹¹

The construction industry depends heavily on contractual obligations of

SELECTED WRITINGS 175-87 (David McLellan ed., Oxford Univ. Press 2004) (1977) (arguing that mode of production creates ideology and then top-down ideology holds society's current mode of production in check).

5. See Mark Pettit, Jr., *Freedom, Freedom of Contract, and the "Rise and Fall,"* 79 B.U. L. REV. 263, 264-66 (1999) ("The idea of contract itself has become more objectified; it is less often an obligation voluntarily assumed by the contracting parties, and more often an obligation imposed by courts to protect the reasonable expectations of others.").

6. Friedman won the Nobel Memorial Prize in Economic Science in 1976, taught at the University of Chicago from 1948 to 1977, contributed to the rise of the Chicago School of Economics, and was "one of the 20th century's leading economic scholars, on a par with giants like John Maynard Keynes and Paul Samuelson." Holcomb B. Noble, *Milton Friedman, the Champion of Free Markets, Is Dead at 94*, N.Y. TIMES, Nov. 17, 2006, at A1.

Galbraith was "one of the most widely read authors in the history of economics." Holcomb B. Noble & Douglas Martin, *John Kenneth Galbraith, 97, Dies; Economist Held a Mirror to Society*, N.Y. TIMES, Apr. 30, 2006, at A1. Though he never won the Nobel Memorial Prize, he published thirty-three books, wrote speeches for President Franklin D. Roosevelt, and advised Presidents John F. Kennedy and Lyndon B. Johnson. *Id.*

7. See Noble, *supra* note 6 ("In forums . . . [Friedman] would spar over the role of government with his more liberal adversaries, including John Kenneth Galbraith . . .").

8. See MILTON FRIEDMAN, AN ECONOMIST'S PROTEST: COLUMNS IN POLITICAL ECONOMY 203 (1972) (arguing that "collectivism is the road to tyranny, inequality, and misery; and that a free market is the only feasible road to freedom and plenty").

9. See JOHN KENNETH GALBRAITH, THE NEW INDUSTRIAL STATE 32-33 (3d ed. 1978) ("The fully planned economy, so far from being unpopular with avowed friends of free enterprise, is warmly regarded by those who know it best.").

10. See Friedrich Kessler, *Contracts of Adhesion—Some Thoughts About Freedom of Contract*, 43 COLUM. L. REV. 629, 629-30 (1943) (arguing that courts must allow business entities to contract freely because it is impossible to predict the limitless number of potential arrangements business entities will need).

11. See CONSTRUCTION LAW 17-32 (William Allensworth et al. eds., 2009) (describing the many relationships formed in a construction project to accomplish the project goals).

parties.¹² Building construction can be a complicated endeavor whether the project is a small apartment complex or a large sports stadium.¹³ One of the many considerations construction project participants must account for is compliance with accessibility guidelines targeting discrimination against persons with disabilities.¹⁴ Regardless of a project's location, most commercial structures must comply with one of two, or possibly both, federal statutes: the Federal Fair Housing Amendments Act of 1988¹⁵ (FHA) and the Americans with Disabilities Act of 1990¹⁶ (ADA) (collectively, "the Acts"). In general, both Acts provide that certain accessibility requirements must be included in new construction projects or modifications to existing structures.¹⁷ Both Acts make failure to comply with accessibility requirements in the "design and construction" of such facilities unlawful discrimination and therefore subject to various remedial actions.¹⁸

Despite a lack of specific language in the FHA and ADA prohibiting indemnification between parties involved in the design and construction of buildings, a few recent federal district court cases have interpreted the Acts to deny implied and express claims for indemnification.¹⁹ Such an interpretation severely limits contractual risk allocation and is not the proper interpretation of the FHA and ADA. To date, no United States circuit court has weighed in on the issue.

This Note traces and critiques the development of the recent trend toward interpreting the FHA and ADA to prohibit contractual indemnification. Part I provides background material about the structure of the construction industry,

12. Eric A. Berg & Bill Hecker, *Accessibility Laws—An Ounce of Prevention Is Worth a Pound of Cure*, 28 CONSTR. LAW. 5, 7 (2008) ("In many relationships directly or tangentially related to the construction industry, parties regularly assign risks to the parties in the best position to guard against them. For example, architects are contractually given responsibility for designing plans and specifications and code compliance").

13. See CONSTRUCTION LAW, *supra* note 11, at 17 ("The financial, technical, business, and regulatory challenges involved in even a small commercial project demand the participation of many diverse participants.").

14. Berg & Hecker, *supra* note 12, at 5.

15. Pub. L. No. 100-430, 102 Stat. 1619 (1988) (codified in scattered sections of 42 U.S.C.).

16. 42 U.S.C. §§ 12101-12213 (2006 & Supp. 2009).

17. See *id.* §§ 3604(f)(3)(C), 12183(a)(1).

18. See *id.*

19. See *United States v. Murphy Dev., LLC*, No. 3:08-0960, 2009 U.S. Dist. LEXIS 100149, at *7-9 (M.D. Tenn. Oct. 27, 2009); *Sentell v. RPM Mgmt. Co.*, No. 4:08CV00629, 2009 U.S. Dist. LEXIS 75126, at *12-13 (E.D. Ark. Aug. 24, 2009); *Mathis v. United Homes, LLC*, 607 F. Supp. 2d 411, 421-23 (E.D.N.Y. 2009); *Equal Rights Ctr. v. Archstone Smith Trust*, 603 F. Supp. 2d 814, 821-22 (D. Md. 2009); *United States v. Shanrie Co.*, 610 F. Supp. 2d 958, 960-61 (S.D. Ill. 2009); *United States v. Gambone Bros. Dev. Co.*, No. 06-1386, 2008 U.S. Dist. LEXIS 73503, at *16-23 (E.D. Penn. Sept. 25, 2008); *Access 4 All, Inc. v. Trump Int'l Hotel & Tower Condo.*, No. 04-CV-7497, 2007 U.S. Dist. LEXIS 13560, at *19-21 (S.D.N.Y. Feb. 26, 2007); *United States v. Quality Built Constr., Inc.*, 309 F. Supp. 2d 767, 778-79 (E.D.N.C. 2003).

indemnification, the FHAA, and the ADA. Part II analyzes and critiques recent cases leading to the potential ban on indemnification clauses for violation of the ADA and FHAA. Part III critiques the recent statutory interpretation prohibiting contractual indemnification and argues that traditional principles of contract law provide a more efficient and effective allocation of responsibility among parties to a construction project, which ultimately benefits not only persons with disabilities, but also society as a whole.

I. BACKGROUND MATERIAL

A. The Construction Industry: A Contractual Nexus of Parties

Construction projects are complex endeavors.²⁰ A thorough analysis of possible combinations of parties and governing relationships is well beyond the scope of this Note. For purposes of examining the impact of contractual indemnification prohibition, familiarity with the basic structure of construction agreements is necessary.

1. *Parties.*—In general, there are three major players involved in designing and constructing covered multifamily dwellings²¹ and commercial facilities: owner, designer, and builder.²² Owners decide a facility is in demand, obtain financing and property, and have an ownership interest in the project.²³ Owners may be public or private entities.²⁴ Designers are licensed architects and engineers who draw plans for projects based on needs and desires of owners, industry standards, and legal requirements.²⁵ Designers also oversee much of the construction project to ensure plans are correctly implemented.²⁶ Builders are usually general contractors and subcontractors.²⁷ Builders coordinate physical construction of the facility by implementing and often adjusting the designer's plans.²⁸

2. *Organization.*—The relationship of owners, designers, and builders may be organized in a number of ways, but most follow the traditional "design-bid-

20. CONSTRUCTION LAW, *supra* note 11, at 17.

21. Not all multifamily buildings are subject to the FHAA. The statute refers to "covered multifamily dwellings," which are: "(A) buildings consisting of 4 or more units if such buildings have one or more elevators; and (B) ground floor units in other buildings consisting of 4 or more units." 42 U.S.C. § 3604(f)(7).

22. CONSTRUCTION LAW, *supra* note 11, at 18-24.

23. *Id.* at 18-20.

24. *Id.* at 18.

25. *Id.* at 20-21.

26. See FUNDAMENTALS OF CONSTRUCTION LAW 10-17 (Carina Y. Enhada et al. eds., 2001). The authors note that the role of designer as project overseer is often overlooked considering its importance in the success of the project. *Id.* "[O]wners often view the architect as their representatives and protectors through the construction process." *Id.* at 8.

27. CONSTRUCTION LAW, *supra* note 11, at 22-25.

28. *Id.*

build” relationship or the “design-build” relationship.²⁹ In design-bid-build relationships, owner, designer, and builder are separate entities.³⁰ Owners enter into contracts with designers to create plans and oversee the construction project.³¹ Once plans are complete, owners hire general contractors to manage the facility’s construction.³² Often, general contractors perform part of the required work and hire subcontractors to perform other requirements of the building project.³³ By contrast, in the design-build context, owners contract with one other party who simultaneously acts as both designer and builder.³⁴ In the design-build scenario, design firms and builder firms may form a joint venture or one firm may offer both services.³⁵

The prevalence of a multi-entity structure where owners, designers, and builders perform distinct functions indicates that it is generally more cost-efficient for the owner to pay designers and builders outside of the owner’s firm rather than bringing them in-house.³⁶ Because architects have special expertise and familiarity with local and national building guidelines,³⁷ it would be very expensive for owners to bring architects in-house if projects are being constructed in many different areas or the owner is developing few projects.³⁸ Just as it is more cost-efficient for most owners to hire outside designers, owners are also likely to hire a builder from outside the firm to make the design a reality because builders possess specialized expertise and equipment most owners lack.³⁹

3. *Relevance of Structure.*—Hypothetically, if an owner employs designers and builders in-house, all liability for failure to design and construct buildings to FHAA and ADA specifications must rest with the owner because it is the only entity involved in design and construction.⁴⁰ But the design and construction of

29. A.H. Gaede, Jr., *Risk Management in Design-Build*, in HANDLING CONSTRUCTION RISKS: ALLOCATE NOW OR LITIGATE LATER 341, 341-46 (2002).

30. FUNDAMENTALS OF CONSTRUCTION LAW, *supra* note 26, at 83.

31. *Id.* at 8-9.

32. *Id.* at 83-84.

33. *Id.* at 83.

34. Gaede, *supra* note 29, at 343.

35. *Id.*

36. See R.H. Coase, *The Nature of the Firm*, 4 ECONOMICA 386, 395 (1937) (“[A] firm will tend to expand until the costs of organising [sic] an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market or the costs of organising [sic] in another firm.”); see also WILLIAM A. KLEIN & JOHN C. COFFEE, JR., BUSINESS ORGANIZATION AND FINANCE: LEGAL AND ECONOMIC PRINCIPLES 19-21 (10th ed. 2007) (comparing organization within firms to organization across markets).

37. See CONSTRUCTION LAW, *supra* note 11, at 132-40.

38. See FUNDAMENTALS OF CONSTRUCTION LAW, *supra* note 26, at 8 (“Many owners are engaged in a onetime or sometime event that is far removed from their primary areas of focus. They do not have the resources to have design or construction expertise in-house.”).

39. *Id.*

40. See JAY M. FEINMAN, PROFESSIONAL LIABILITY TO THIRD PARTIES 3-6 (2d ed. 2007) (noting that in order to have third party professional liability, there must be a contractual

buildings is complex and almost always a multi-entity affair.⁴¹ Because the owner owns the building, it is the most likely target for grievances associated with defects in the building.⁴² Yet the owner hired the designer and builder from outside its firm in part because it did not possess the expertise needed to design and construct the building.⁴³ In response to the risk of lawsuit, owners often pay premiums for express indemnification clauses covering aspects of the project in which the owner has limited expertise and control.⁴⁴ The accompanying shift in liability gives the other parties an incentive to minimize the probability of payout by performing in a risk-minimizing fashion.⁴⁵

B. Indemnification

Indemnification is a risk-allocation tool that is widely used in business negotiations and commonplace in the construction industry.⁴⁶ Indemnification can be defined as a “complete shifting of liability for loss from one party to another. In essence, one person either agrees or is compelled by law to hold another person harmless for loss or damage which the second person has or is anticipated to sustain because of some liability to a third person.”⁴⁷ Indemnification is also derivative in nature.⁴⁸ One party must be found liable to a third party before the liable party may seek indemnification.⁴⁹ Indemnification may be express or implied by law.⁵⁰

Implied indemnification is court-compelled liability shifting for public policy reasons or because the nature of the relationship between parties indicates an implied agreement to indemnify.⁵¹ The employer-employee relationship is one

relationship between at least two parties).

41. CONSTRUCTION LAW, *supra* note 11, at 17 (“The design and construction of a project is a collaborative process requiring talent, execution, and coordination of many different people and organizations. As the size, cost, complexity, or unusual features of a project increase, the number of participants . . . likely will increase as well.”).

42. See FUNDAMENTALS OF CONSTRUCTION LAW, *supra* note 26, at 10-17.

43. See *supra* note 38 and accompanying text.

44. See Kenneth M. Cushman & Joyce K. Hackenbrach, *Construction Project Risk Allocation: The Owner's Perspective*, in HANDLING CONSTRUCTION RISKS: ALLOCATE NOW OR LITIGATE LATER, *supra* note 29, at 9, 10-12.

45. See *id.* at 9-14.

46. THE CONSTRUCTION CONTRACTS BOOK: HOW TO FIND COMMON GROUND IN NEGOTIATING THE 2007 INDUSTRY FORM CONTRACT DOCUMENTS 29 (Daniel S. Brennan et al. eds., 2d ed. 2008).

47. Bruce H. Schoumacher, *Risk Management and Indemnity* § 4-13, in CONSTRUCTION LAW 13.17 (Steven G.M. Stein ed., 2010).

48. *Id.* at § 1.

49. *Id.*

50. *Id.*

51. *Id.*

common type of arrangement that indicates an implied agreement to indemnify.⁵² Public policy may demand indemnification if a plaintiff succeeds against one defendant while a more culpable defendant escapes liability.⁵³ Implied indemnification is essentially an extension of vicarious liability principles in tort law.⁵⁴

More important for purposes of this Note, express indemnification is based on contract rather than tort law.⁵⁵ As a creature of contract, an express indemnification agreement evidences the parties' intent to shift risks *ex ante*.⁵⁶ Predictability is one of the great advantages of contract law generally, and express indemnification specifically provides predictability.⁵⁷ Presumably, bargaining parties pay and receive premiums reflected in the contract price according to their exposure to liability for damages on the occurrence of some event.⁵⁸ As the potential for payout and forecasted damages increases, the premium paid for protection against risk likewise increases.⁵⁹

Express indemnification is a widely accepted vehicle for shifting risk in business negotiations, but there are a few instances in which courts and legislatures deny enforceability of indemnification provisions.⁶⁰ Parties generally cannot seek indemnification against their own willful harm-causing conduct.⁶¹ Additionally, parties generally cannot seek indemnification against conduct they know to be illegal or immoral at the time of contracting.⁶² For instance, parties may not agree to indemnify others for damages resulting from fraud.⁶³ Negligent behavior, as opposed to willful behavior, is more questionable.⁶⁴ Generally, parties are able to contractually indemnify against their own negligence, but several states have enacted anti-indemnification statutes that limit the ability of parties to seek indemnification for their own negligence.⁶⁵ Anti-indemnification statutes may limit or prohibit indemnification in situations where legislatures are concerned that shifting liability may lead to immoral, inefficient, or negligent

52. *Id.* at § 1(a).

53. *Id.*

54. *Id.*

55. *Id.* at § 1(b).

56. THE CONSTRUCTION CONTRACTS BOOK, *supra* note 46, at 29.

57. See ANTHONY T. KRONMAN & RICHARD A. POSNER, THE ECONOMICS OF CONTRACT LAW 4 (1979) ("An important function of contract law is to enforce the parties' agreed-upon allocation of risk.").

58. See RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 107 (4th ed. 1992) (discussing contract modification and noting that "[i]f . . . [the risk] was allocated to the crew they were presumably compensated for assuming it").

59. See *id.*

60. Schoumacher, *supra* note 47, at § 1(b)(ii).

61. *Id.*

62. *Id.*

63. *Id.* at § 1(b)(iii).

64. *Id.* at § 1(b)(ii).

65. *Id.* at § 2.

behavior.⁶⁶

In the construction industry, bargaining is essential to the process of allocating risk and responsibility to parties in the best position to minimize such risks.⁶⁷ As evidence of the wide acceptance of indemnification in the construction industry, a recent article explored indemnification among architects, contractors, and developers under the ADA and FHAA.⁶⁸ The authors noted that because the Acts are a hybrid of building code and civil rights legislation, courts have been reluctant to allow complete abdication of liability by contractors, architects, or developers when a structure fails to comply with ADA and FHAA accessibility guidelines.⁶⁹ Yet the authors point out that indemnification may still be accomplished through the private contractual relationship of the parties.⁷⁰ The authors provide an example of a contractual provision between an owner and contractor:

Owner acknowledges that Contractor has no design responsibility hereunder. To the extent permitted by law, if Contractor is ever named as a defendant in a lawsuit brought pursuant to, or held liable for violation of, any federal, state, or local disabled-access statute, including [but] not limited to the Americans with Disabilities Act, the Owner agrees to indemnify, defend, and hold harmless the Contractor for any and all liability thereunder, including but not limited to fines, judgments, costs, attorney fees, and expert witness fees.⁷¹

The authors note that “contractual provisions such as these will likely not lead to an architect or contractor being dismissed from an ADA lawsuit as a matter of law. However, they will shift the ultimate responsibility for paying to remediate noncompliant structures (and the attendant legal costs).”⁷² This conclusion was sensible in 2008 when the article was published, but recent district court interpretations of the FHAA and ADA indicate a movement toward unenforceability of these indemnification provisions.⁷³

66. *Id.* at § 1(b)(ii); *see, e.g.*, ARIZ. REV. STAT. ANN. § 34-226 (2010) (construction contracts indemnifying against promisee’s own negligence void); CONN. GEN. STAT. ANN. § 52-572k (West, Westlaw through 2010 legislation) (similar); GA. CODE ANN. § 13-8-2 (2010) (similar); MONT. CODE ANN. § 28-11-302 (2009) (“An agreement to indemnify a person against an act thereafter to be done is void if the act be known by such person, at the time of doing it, to be unlawful.”); TENN. CODE ANN. § 62-6-123 (West, Westlaw through 2010 legislation) (similar to Arizona).

67. *See* Cushman & Hackenbrach, *supra* note 44, at 9-14.

68. Berg & Hecker, *supra* note 12, at 5.

69. *Id.* at 5, 7.

70. *Id.* at 7.

71. *Id.* at 8.

72. *Id.*

73. *See supra* text accompanying note 19.

C. The FHAA

In order to better evaluate the movement toward indemnification prohibition, some background about each Act is useful for perspective. In 1988, Congress passed the FHAA to amend the Fair Housing Act of 1968 (FHA).⁷⁴ The original FHA aimed to curb discrimination in housing but failed to address housing practices that disadvantaged persons with disabilities.⁷⁵ As part of the 1988 amendments, Congress added provisions targeting discrimination against persons with disabilities in the housing market.⁷⁶ Congress realized that disability-based discrimination was different from race, gender, nationality, or religious-based discrimination.⁷⁷ In the latter examples, the physical attributes of dwellings are not relevant for determining whether an individual or entity has discriminated in rental or sale.⁷⁸ Disabilities pose a unique problem because it is possible to discriminate strictly on the basis of designing and constructing dwellings with features that make them unusable by persons with disabilities.⁷⁹

In order to address these unique physical design problems, the FHAA sets out seven requirements for those who “design and construct” post-enactment “covered multifamily dwellings.”⁸⁰ The requirements are that

- (i) the public use and common use portions of such dwellings are readily accessible to and usable by handicapped persons;
- (ii) all the doors designed to allow passage into and within all premises within such dwellings are sufficiently wide to allow passage by handicapped persons in wheelchairs; and
- (iii) all premises within such dwellings contain the following features of adaptive design:
 - (I) an accessible route into and through the dwelling;
 - (II) light switches, electrical outlets, thermostats, and other environmental controls in accessible locations;
 - (III) reinforcements in bathroom walls to allow later installation of grab bars; and
 - (IV) usable kitchens and bathrooms such that an individual in a wheelchair can maneuver about the space.⁸¹

Construction project participants have engaged in litigation to determine whether the FHAA’s “design and construct” language applies to owners, designers, and builders in their distinct roles or if the language only targets

74. Pub. L. No. 100-430, 102 Stat. 1619 (1988) (codified as amended at 42 U.S.C. §§ 3601-3619 (2006)).

75. H.R. REP. NO. 100-711, at 13 (1988), *reprinted in* 1988 U.S.C.C.A.N. 2173, 2174.

76. *Id.*

77. See H.R. REP. NO. 100-711, at 18, *reprinted in* 1988 U.S.C.C.A.N. 2173, 2179.

78. *Id.*

79. *Id.*

80. 42 U.S.C. § 3604(f)(3)(C) (2006).

81. *Id.*

participants playing dual roles as both designer and builder.⁸² Courts have interpreted the FHAA to make all parties involved in a project potentially liable for failure to comply with FHAA provisions.⁸³ One court has noted that “[d]esign and construct’ is a broad sweep of liability . . . [encompassing] architects, builders, and planners.”⁸⁴ Another important decision took an even broader view by finding that “[w]hen a group of entities enters into the design and construction of a covered dwelling, all participants *in the process as a whole* are bound to follow the FHAA.”⁸⁵

D. The ADA

In 1990, Congress passed the ADA⁸⁶ essentially as an extension of the Civil Rights Acts of 1964⁸⁷ and 1968.⁸⁸ Similar to the FHAA, Congress recognized a need to extend existing protections against discriminatory practices to benefit persons with disabilities.⁸⁹ The ADA is comprehensive legislation covering several discriminatory practices persons with disabilities may face in areas such as employment and transportation services.⁹⁰

Title III of the ADA includes anti-discrimination provisions targeting accessibility in public accommodations.⁹¹ The ADA makes the “failure to design and construct facilities for first occupancy later than 30 months after July 26, 1990, that are readily accessible to and usable by individuals with disabilities” unlawful discrimination.⁹² Unlike the FHAA, the ADA does not list specific

82. See, e.g., Mont. Fair Hous. v. Am. Capital Dev., Inc., 81 F. Supp. 2d 1057, 1062, 1068-69 (D. Mont. 1999) (in which defendant builder argued that it only built and did not design the apartment complexes in FHAA case).

83. See, e.g., Balt. Neighborhoods, Inc. v. Rommel Builders, Inc., 3 F. Supp. 2d 661, 664 (D. Md. 1998) (“Defendant’s narrow interpretation of the ‘design and construct’ provision would defeat the purpose of the FHAA by allowing architects and builders who are involved in either the design or construction, but not both, to escape liability”).

84. United States v. Days Inn of Am., Inc., 997 F. Supp. 1080, 1083 (C.D. Ill. 1998), *aff’d*, 151 F.3d 822 (8th Cir. 1998).

85. *Balt. Neighborhoods*, 3 F. Supp. 2d at 665 (emphasis in original).

86. 42 U.S.C. §§ 12101-12213.

87. Pub. L. No. 88-352, 78 Stat. 241 (codified as amended at 42 U.S.C. §§ 2000(a)-2000(h)(6)).

88. Pub. L. No. 90-284, 82 Stat. 73 (codified as amended at 42 U.S.C. §§ 3601-3619, 3631); see H.R. REP. NO. 101-485(III), at 26 (1990), reprinted in 1990 U.S.C.C.A.N. 445, 449.

89. See H.R. REP. NO. 101-485(III), at 23 (1990), reprinted in 1990 U.S.C.C.A.N. 445, 446.

90. See 42 U.S.C. §§ 12101-12189 (2006 & Supp. 2009).

91. See *id.* § 12182. Places of public accommodation generally include: (1) places of lodging; (2) places that serve food and drink; (3) places of exhibition and entertainment; (4) places of public gathering; (5) sales or rental establishments; (6) service establishments; (7) certain public transportation; (8) places holding a public display or collection; (9) places of recreation; (10) places of education; (11) social service centers; and (12) places of exercise or athletic recreation. *Id.*

92. *Id.* § 12183(a)(1).

design requirements that must be followed in order for places of public accommodation to be ADA-compliant.⁹³ Instead, design requirements are listed in the ADA Accessibility Guidelines.⁹⁴

As with the FHAA, parties have litigated whether the ADA's "design and construct" language makes owners, designers, and builders liable or only parties responsible for both designing and constructing facilities.⁹⁵ Interestingly, the argument that designers or builders are not liable for noncompliance is stronger under the ADA because while the "design and construct" section does not clearly list proper potential defendants⁹⁶ (the same as the FHAA), the immediately preceding section only targets "any person who owns, leases (or leases to), or operates a place of public accommodations."⁹⁷ While some courts have accepted this argument,⁹⁸ most have followed the FHAA interpretation by holding all parties involved in a project potentially liable for failure to comply with the provisions of the ADA.⁹⁹

II. THE ROAD TO FHAA AND ADA CONTRACTUAL INDEMNIFICATION PROHIBITION

Even though Congress enacted the FHAA and ADA in 1988 and 1990, respectively,¹⁰⁰ and both Acts are ambiguous regarding whether indemnification is prohibited,¹⁰¹ no court interpreted either Act regarding indemnification until 2003.¹⁰² After 2003, there was a five-year dearth of interpretive cases, but since 2008, six federal district courts have interpreted the Acts to prohibit contractual indemnification in design and construct cases.¹⁰³ This section first analyzes the 2003 opinion and then considers one of the more recent cases.

A. United States v. Quality Built Construction, Inc.¹⁰⁴

In 2003, *Quality Built* was the first district court case to deny indemnification for failure to design and construct buildings according to FHAA or ADA

93. Compare 42 U.S.C. § 3604(f)(3)(C) (FHAA), with 42 U.S.C. § 12183 (ADA).

94. See 28 C.F.R. § 36 (2009).

95. See, e.g., *United States v. Days Inn of Am., Inc.*, 151 F.3d 822, 824-27 (8th Cir. 1998).

96. See 42 U.S.C. § 12183(a)(1).

97. *Id.* § 12182(a); see also Mita Chatterjee, Case Note, *Access Denied and Not Designed: The Ninth Circuit Drafts a Narrow Escape for Architect Liability Under the Americans with Disabilities Act* in *Lonberg v. Sanborn Theaters, Inc.*, 9 VILL. SPORTS & ENT. L.J. 293, 303-05 (2002).

98. See, e.g., *Lonberg v. Sanborn Theaters, Inc.*, 259 F.3d 1029, 1033-36 (9th Cir. 2001).

99. See, e.g., *Days Inn of Am., Inc.*, 151 F.3d at 824-27.

100. See *supra* Parts I.C-D.

101. See *infra* Part III.A.

102. See *supra* note 19 (listing all cases interpreting either or both Acts).

103. See *supra* note 19 (listing decisions handed down since 2008).

104. 309 F. Supp. 2d 767 (E.D.N.C. 2003).

requirements.¹⁰⁵ Quality Built Construction (“Quality Built”), a North Carolina development company, hired Hite/MSM, P.C. (“Hite”), an architectural firm, to design apartment complexes.¹⁰⁶ The United States brought an action against Quality Built and Hite for violation of numerous design and construction provisions of the FHAA.¹⁰⁷ Quality Built claimed that it had relied upon Hite to adequately design the complexes in accordance with “all applicable codes and building guidelines.”¹⁰⁸ Quality Built filed a cross-claim against Hite seeking indemnification and contribution¹⁰⁹ under the FHAA.¹¹⁰ Hite argued that Quality Built could not seek indemnification because the FHAA does not expressly or impliedly provide for indemnification between co-defendants.¹¹¹

The *Quality Built* court recognized that indemnification under the FHAA was an issue of first impression.¹¹² Rather than interpreting the FHAA to allow indemnification, the court followed reasoning laid down in *Northwest Airlines v. Transport Workers Union of America*,¹¹³ a federal employment discrimination case. Borrowing from *Northwest Airlines*, the *Quality Built* court determined that the FHAA does not provide a right to sue for indemnification because: (1) Congress did not expressly provide such a remedy; (2) Congress did not imply, either through statutory language or legislative history, that such a remedy should be available between co-defendants; and (3) Congress did not provide courts with the ability to fashion federal common law in this area.¹¹⁴

Although Quality Built also brought state law cross-claims for breach of contract and breach of the standard of care against Hite, the court did not squarely address the nature of the claims or give a thorough analysis of potential

105. Robert G. Schwemm, *Barriers to Accessible Housing: Enforcement Issues in “Design and Construction” Cases Under the Fair Housing Act*, 40 U. RICH. L. REV. 753, 811 (2006).

106. *Quality Built*, 309 F. Supp. 2d at 771.

107. *Id.*

108. *Id.* at 778.

109. As opposed to indemnification, contribution shifts only a portion of the costs of liability rather than all of the costs from one party to another. See *Baker, Watts & Co. v. Miles & Stockbridge*, 876 F.2d 1101, 1103 (4th Cir. 1989). Potential contribution prohibition under the Acts is beyond the scope of this Note. Because this Note focuses primarily on express indemnification, further explanation of *Quality Built*’s treatment of contribution has been omitted.

As to the nature of Quality Built’s indemnification claim, the court’s opinion lacks clarity. Certainly Quality Built was arguing that it should be permitted to seek indemnification on implied grounds through a cross-claim. See *Quality Built*, 309 F. Supp. 2d at 778. Quality Built also brought state law breach of contract actions that were challenged by Hite as de facto claims for indemnification. *Id.* This may indicate that Quality Built and Hite had an express indemnification agreement in their contract documents. Unfortunately, the court does not address whether there were such contractual provisions, and party briefs are unavailable.

110. *Quality Built*, 309 F. Supp. 2d at 778.

111. *Id.*

112. *See id.*

113. 451 U.S. 77 (1981). For further discussion of *Northwest Airlines*, see *infra* Part II.B.1.

114. *Quality Built*, 309 F. Supp. 2d at 778-79.

contractual indemnification.¹¹⁵ With respect to the state law claims, the court held that “[t]o the extent that Defendants seek indemnification on the basis of these state actions, the claims are not allowed.”¹¹⁶ The court did not address the nature of the parties’ contract documents.¹¹⁷

B. Choosing Precedent for Interpreting the FHAA and ADA

The *Quality Built* court was the first court to interpret the FHAA or ADA to preclude indemnification between construction project parties.¹¹⁸ As courts often do, the *Quality Built* court structured its reasoning by claiming to slavishly follow precedent without laying out potential alternatives. Despite the persuasive tone of the *Quality Built* opinion, *Northwest Airlines* was not mandatory authority because it was not directly on point¹¹⁹ and the court had other available options.¹²⁰

1. *Problems with Following Northwest Airlines.*—The *Quality Built* court’s chosen precedent was problematic because it was an employment discrimination case.¹²¹ In *Northwest Airlines*, a union discriminated against potential job candidates in its placement program with an employer.¹²² The plaintiff sued the employer, who was held liable for discriminatory hiring practices under federal Title VII¹²³ anti-employment discrimination provisions and the Equal Pay Act.¹²⁴ The employer filed a claim against the union for contribution.¹²⁵ While courts interpreting the FHAA often look to federal employment discrimination cases for guidance,¹²⁶ the *Quality Built* court failed to address important differences between employment discrimination and design and construct cases.

Like the FHAA and ADA, Title VII neither expressly prohibits nor expressly allows contribution or indemnification.¹²⁷ The *Northwest Airlines* Court interpreted Title VII to prohibit contribution.¹²⁸ Prohibiting contribution in Title

115. *See id.*

116. *Id.* at 779.

117. *Id.*

118. For a list of cases with dates of decisions, see *supra* note 19.

119. *See Quality Built*, 309 F. Supp. 2d at 778 (“The [*Quality Built*] [c]ourt has not found any cases directly on point.”).

120. *See Schwemm, supra* note 105, at 807-13. This point is elaborated further in Part II.B.2.

121. *See Nw. Airlines v. Transp. Workers Union*, 451 U.S. 77, 79-82 (1981).

122. *Id.*

123. Title VII of the Civil Rights Act of 1964 prohibits employers from discriminating against employees based on “race, color, religion, sex, or national origin.” 42 U.S.C. § 2000e-2(a)(1) (2006).

124. The Equal Pay Act of 1963 generally prohibits employers from discriminatorily paying employees of the opposite sex disparate wages for tasks that require “equal skill, effort, and responsibility, and which are performed under similar working conditions.” 29 U.S.C. § 206(d).

125. *Nw. Airlines*, 451 U.S. at 79-82.

126. *Schwemm, supra* note 105, at 776-77.

127. *Nw. Airlines*, 451 U.S. at 91.

128. *Id.* at 98.

VII cases makes sense, but employment discrimination is significantly different from failure to design and construct compliant buildings under the FHAA and ADA.

To begin with, the context of making hiring and firing decisions is different from making accessibility design decisions. As opposed to construction projects, employers making hiring decisions rarely rely on indemnification to shift risk because employers do not heavily rely on the expertise of outside parties to make employment decisions.¹²⁹ Most cases of employment discrimination are comparable to owners having designers and builders in-house—that is, only one party is potentially liable for discrimination.¹³⁰ Not only do employers have the necessary expertise to choose job candidates, but they are presumably more familiar with their own needs than an outside party such as a union.¹³¹ By contrast, indemnification is prevalent in the construction industry precisely because outside parties have more expertise and familiarity than the party seeking indemnification.¹³²

Second, the statutory elements needed to make a claim in employment discrimination cases are different from accessibility design and construction cases. Employment discrimination claims under Title VII generally focus on discriminatory intent.¹³³ Noncompliant facilities under the FHAA and ADA give rise to liability regardless of whether a plaintiff can show discriminatory intent.¹³⁴ As discussed above, parties generally cannot indemnify against intentionally unlawful behavior.¹³⁵ Logically, it makes no sense for Congress to allow indemnification in employment discrimination because that would allow parties to shift liability for intentional wrongdoing. By contrast, when parties contractually shift liability for FHAA and ADA accessibility design compliance, they are aligning responsibility with expertise rather than indemnifying against intentionally wrongful behavior.¹³⁶ Even if an owner intentionally disregards the

129. See Elaine W. Shoben, *Employee Recruitment by Design or Default: Uncertainty Under Title VII*, 47 OHIO ST. L.J. 891, 904-07 (1986) (describing the use of outside parties in employment hiring decisions as “restricted recruitment” and acknowledging that employers remain the parties with most knowledge and responsibility for recruitment needs).

130. See *supra* Part I.A.

131. See Shoben, *supra* note 131, at 904-07.

132. See *supra* Part I.B.

133. Under Title VII, disparate impact claims provide a type of strict liability without intent, but disparate impact cases only provide a presumption of intent that can be rebutted by the employer. See *Espinoza v. Farah Mfg. Co.*, 414 U.S. 86, 96-97 (1973). Design and construct cases are all essentially disparate impact claims, but the disparate treatment yields liability rather than a rebuttable presumption. See, e.g., *United States v. Pac. Nw. Elec., Inc.*, No. CV-01-019-S-BLW, 2003 U.S. Dist. LEXIS 7990 at *47-48 (D. Idaho 2003) (FHAA); *Indep. Living Res. v. Or. Arena Corp.*, 982 F. Supp. 698, 707 (D. Or. 1997) (ADA).

134. See, e.g., *Pac. Nw. Elec., Inc.*, 2003 U.S. Dist. LEXIS 7990 at *47-48 (FHAA); *Or. Arena Corp.*, 982 F. Supp. at 707 (ADA).

135. See *supra* Part I.B.

136. See *supra* Part I.B. & *infra* Part III.B.2.b.

mandates of the FHAA and ADA, an indemnification clause is traditionally unenforceable.¹³⁷

Third, in design and construct cases the motives for noncompliance are different from the motives of employers that violate Title VII. As Congress recognized, failure to comply with FHAA and ADA requirements in design and construct cases is probably the result of ignorance rather than animosity.¹³⁸ By contrast, social stereotypes drive discrimination in employment decisionmaking rather than ignorance.¹³⁹ Moreover, because employers are generally the only hiring and firing decisionmakers, knowledge of discrimination guidelines is within their field of expertise.¹⁴⁰ There is therefore no reason for employers to seek indemnification as a means of combating ignorance. In the construction industry, indemnification is a useful tool for combating ignorance because it creates an added incentive for the party in the best position to avoid the cost of liability to ensure that facilities are compliant.¹⁴¹

2. *An Alternative Path.*—Professor Robert Schwemm¹⁴² recognized an alternative path that *Quality Built* could have followed instead of adopting the *Northwest Airlines* approach. Rather than foreclosing traditional principles of tort law that would impose joint and several liability, *Quality Built* could have followed the reasoning of two Supreme Court decisions: *Meyer v. Holley*¹⁴³ and *Norfolk & Western Railway Co. v. Ayers*.¹⁴⁴ In *Meyer*, the Court reaffirmed that

137. See *supra* Part I.B. (indemnification for intentional wrongdoing generally prohibited).

138. See H.R. REP. NO. 100-711, at 25, reprinted in 1988 U.S.C.C.A.N. 2173, 2186 (approvingly quoting the Supreme Court as observing that discrimination on the basis of disability is “most often the product, not of invidious animus, but rather of thoughtlessness and indifference—of benign neglect” (quoting *Alexander v. Choate*, 469 U.S. 287, 297 (1985))). A desire to cut costs is another potential motive, but according to Congress, additional costs should be negligible. See H.R. REP. NO. 100-711, at 80.

139. See D. Wendy Greene, *Title VII: What’s Hair (and Other Race-Based Characteristics) Got to Do with It?*, 79 U. COLO. L. REV. 1355, 1376-84 (2008) (analyzing and applying the “cultural meaning” test and unconscious racism in the specific context of employer decisionmaking) (citing Charles R. Lawrence III, *The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism*, 39 STAN. L. REV. 317 (1987) and R.A. Lenhardt, *Understanding the Mark: Race, Stigma, and Equality in Context*, 79 N.Y.U.L.REV. 803 (2004)).

140. See GEORGE RUTHERGLEN, *EMPLOYMENT DISCRIMINATION LAW: VISIONS OF EQUALITY IN THEORY AND DOCTRINE* 31-36 (2d ed. 2007).

141. See *supra* Part I.B. & *infra* Part III.B.2.b.

142. Professor Schwemm is the Ashland Professor of Law at the University of Kentucky College of Law and has authored several articles addressing housing discrimination. E.g., Eugene R. Gaetke & Robert G. Schwemm, *Government Lawyers and Their Private “Clients” Under the Fair Housing Act*, 65 GEO. WASH. INT’L L. REV. 329 (1997); Robert G. Schwemm, *Discriminatory Housing Statements and 3604(c): A New Look at the Fair Housing Act’s Most Intriguing Provision*, 29 FORDHAM URB. L.J. 187 (2001); Robert G. Schwemm & Michael Allen, *For the Rest of Their Lives: Seniors and the Fair Housing Act*, 90 IOWA L. REV. 121 (2004).

143. 537 U.S. 280 (2003).

144. 538 U.S. 135 (2003); see Schwemm, *supra* note 105, at 807-13 (recognizing *Norfolk* as

because the FHAA essentially creates tort liability, traditional tort principles should apply unless Congress explicitly foreclosed them.¹⁴⁵ In *Norfolk*, a Federal Employers' Liability Act¹⁴⁶ (FELA) case, the Court interpreted the FELA using the same method of construction as in *Meyer* and held that the Act did not foreclose contribution¹⁴⁷ because “joint and several liability is the traditional rule.”¹⁴⁸ Had the *Quality Built* court followed this approach, it would likely have allowed contribution claims, but it might still have foreclosed implied indemnification claims on grounds that implied indemnification is not necessarily a favored traditional tort principle.¹⁴⁹ Even so, prohibition of implied indemnification is radically different from prohibition of express contractual indemnification.¹⁵⁰

As opposed to claims for implied indemnification, express contractual indemnification is a traditionally accepted vehicle for shifting liability in many tort situations.¹⁵¹ Express contractual indemnification is treated differently because it represents an important recognition between the parties *ex ante* that a particular party should bear all responsibility for a task.¹⁵² In this way, contractual indemnification functions as an efficiency mechanism.¹⁵³ By contrast, claims for contribution and implied indemnification consider fault and responsibility *ex post*.¹⁵⁴ The *Quality Built* court failed to distinguish contribution and implied indemnification from express indemnification. As a result, the court also failed to consider whether Congress intended to eliminate such a traditionally useful construction industry tool for allocating responsibility *ex ante* and properly

a potential alternative precedent).

145. *Meyer*, 537 U.S. at 828-29 (citing *Curtis v. Loether*, 415 U.S. 189, 195-96 (1974) and *United States v. Texas*, 507 U.S. 529, 534 (1993) (“In order to abrogate a common-law principle, the statute must ‘speak directly’ to the question addressed by the common law.”)).

146. 45 U.S.C. §§ 51-60 (2006).

147. As opposed to indemnification, contribution shifts only a portion of the costs of liability rather than all of the costs from one party to another. *See United States v. Quality Built Constr., Inc.*, 309 F. Supp. 2d 767, 778-80 (E.D.N.C. 2003).

148. *Norfolk*, 538 U.S. at 159-66 (quoted language at 163).

149. “[Implied] [i]ndemnity was and is permitted in only a few situations [in the joint and several liability system].” Schwemm, *supra* note 105, at 811 (citing DAN D. DOBBS, THE LAW OF TORTS 1079 (2002)).

150. Again, it is unclear whether *Quality Built* stands for the proposition that express contractual indemnification should be prohibited. The court failed to squarely address the issue. *See supra* note 111 and second paragraph of accompanying text.

151. While strict liability in tort and intentional torts involve acts that parties generally cannot indemnify against, liability arising from the most common tort claim—negligence—can generally be shifted through express contractual indemnification. *See supra* Part I.B.

152. *See POSNER, supra* note 58, at 119 (“[A]n important function of contracts is to assign risk to superior risk bearers,” and “[i]f the risk materializes, the party to whom it was assigned must pay.”).

153. *Id.*

154. Schoumacher, *supra* note 47.

incentivizing parties.

3. *Unresolved Issues*.—*Quality Built* focused on the FHAA rather than the ADA,¹⁵⁵ leaving unresolved whether the ADA should be treated differently.¹⁵⁶ Additionally, *Quality Built* failed to squarely address express contractual indemnification,¹⁵⁷ resulting in another unresolved issue. Further, *Quality Built* did not analyze *Meyer* and *Norfolk*—an avenue left open to future district courts.¹⁵⁸

After a lapse of five years, several other federal courts addressed indemnification in FHAA and ADA design and construct suits.¹⁵⁹ None of these courts challenged the reasoning of *Quality Built* or addressed differences between employment discrimination and design and construct cases. Perhaps the district courts have construed Congress's inactivity since the *Quality Built* decision as congressional affirmation of indemnification prohibition. On the other hand, Congress does not always respond to court decisions—especially in the case of a single district court. Also, congressional inactivity after *Quality Built* only indicates that Congress may agree that the FHAA does not allow co-defendants to file cross-claims for implied indemnification. Because the *Quality Built* court failed to squarely address express contractual indemnification, Congress's silence does not necessarily indicate approval of express indemnification prohibition. As discussed further below, district courts that have addressed contractual indemnification have apparently followed *Quality Built* because they are convinced that by depriving parties of the ability to shift risk and costs, all project participants have increased incentive to ensure compliance with the Acts.¹⁶⁰

155. See *United States v. Quality Built Constr., Inc.*, 309 F. Supp. 2d 767, 778-80 (E.D.N.C. 2003).

156. As indicated in Part II.C., recent court decisions have extended express indemnification prohibition to ADA cases without addressing differences in the Acts. Space limitations foreclose a thorough analysis of this topic, but a terse discussion is warranted. Plausibly, express indemnification prohibition stands on more solid footing in FHAA cases than ADA cases. Congress allowed for punitive damages in the FHAA design and construct context, but not in the ADA. Compare 42 U.S.C. § 3613 (2006) (FHAA remedies), with 42 U.S.C. § 12188 (ADA remedies). Congress's imposition of punishment could be taken as an indication that parties should not be able to escape punishment by shifting costs. On the other hand, wise drafters of indemnification clauses can make exceptions for punitive damages. Also, punishment is presumably only warranted in cases of intentional wrongdoing, and in such cases indemnification clauses are void anyway. See *supra* Part II.B.

157. See *Quality Built*, 309 F. Supp. 2d at 778-80.

158. Schwemm, *supra* note 105, at 809.

159. See *supra* text accompanying note 19.

160. See *infra* Part III.D.

C. Equal Rights Center v. Archstone Smith Trust:¹⁶¹ *A Showcase of the Trend in Design and Construct Suits*

Although there have been a number of recent district court decisions following *Quality Built's* lead,¹⁶² *Equal Rights Center v. Archstone Smith Trust*¹⁶³ succinctly evidences the trend toward prohibiting express contractual indemnification under both the FHAA and the ADA with more thorough analysis than other cases.¹⁶⁴

1. *Background.*—On December 20, 2004, the Equal Rights Center, the American Association of People with Disabilities, and the United Spinal Association filed an action against several entities involved in the design, construction, maintenance, and operation of apartment complexes throughout the United States.¹⁶⁵ Among those entities were Archstone Smith Trust and Archstone Operating Trust (collectively “Archstone”), a real estate investment trust specializing in apartment development,¹⁶⁶ and Niles Bolton,¹⁶⁷ a provider of architectural services for fifteen of the properties in question.¹⁶⁸ The plaintiffs alleged violations of both the FHAA and ADA, including failure to design and construct the apartment complexes in compliance in seventeen states and the District of Columbia.¹⁶⁹ By March of 2009, all defendants had reached settlement agreements, but one important issue remained before the court: Archstone’s state law cross-claim against Niles Bolton for indemnification.¹⁷⁰

2. *The Issue.*—Archstone conceded that its apartment complexes were in violation of the FHAA and ADA.¹⁷¹ As a result, it agreed to pay 1.4 million dollars in damages, attorneys’ fees, costs, and other expenses.¹⁷² It also agreed

161. 603 F. Supp. 2d 814 (D. Md. 2009).

162. *See supra* text accompanying note 19.

163. 603 F. Supp. 2d 814.

164. *Id.*

165. *Id.* at 815-16.

166. Archstone is one of the largest real estate investment trusts in the United States. It is a publicly traded corporation and has a long history of mergers and acquisitions. *See About Us, ARCHSTONE APARTMENTS*, http://www.archstoneapartments.com/Top/About_Us.htm (last visited Apr. 4, 2011); *Archstone-Smith Trust*, FUNDING UNIVERSE, <http://www.fundinguniverse.com/company-histories/ArchstoneSmith-Trust-Company-History.html> (last visited Apr. 4, 2011).

167. Niles Bolton is an architectural firm headquartered in Atlanta, Georgia. It has 190 employees and does business throughout the United States as well as internationally. *See NILES BOLTON ASSOCS.*, <http://www.nilesbolton.com/> (last visited Apr. 4, 2011).

168. *Archstone*, 603 F. Supp. 2d at 815-16.

169. The states in question were Arizona, California, Colorado, Florida, Georgia, Illinois, Maryland, Massachusetts, New Mexico, North Carolina, New Jersey, New York, Oregon, Tennessee, Texas, Virginia, and Washington. *Id.* at 816.

170. Archstone originally only brought a claim for indemnification against Niles Bolton and sought to add a claim for contribution during the 2008 appeal. *Id.* at 817.

171. *Id.* at 815-16.

172. *Id.*

to pay to retrofit noncompliant apartment complexes to bring them in line with FHAA and ADA mandates.¹⁷³ Archstone argued that Niles Bolton was obligated to repay Archstone for the portion of damages Niles Bolton caused Archstone to incur specifically because of designs that failed to comply with the FHAA and ADA.¹⁷⁴ Archstone “[sought] damages only for those violations . . . that occurred because . . . [Niles Bolton] specified an incorrect dimension or other detail in its construction documents, or otherwise failed to provide sufficient information for the builder to construct the project in accordance with the applicable accessibility requirements.”¹⁷⁵

Archstone asserted its indemnification claim both on an implied indemnification ground and on the basis of express contractual indemnification.¹⁷⁶ According to contracts between Niles Bolton and Archstone,¹⁷⁷ Niles Bolton was responsible for designing the structures in compliance with all federal laws, including the FHAA and ADA.¹⁷⁸ Additionally, Niles Bolton promised to indemnify Archstone for any costs Archstone incurred as a result of Niles Bolton’s failure to properly design the buildings.¹⁷⁹

3. *Holding and Rationale.*—The *Archstone* court held that Archstone’s implied tort law and express contract law indemnification claims failed, and Niles Bolton did not have to compensate Archstone (though Niles Bolton had to pay whatever its agreed settlement amount was with the original plaintiffs).¹⁸⁰ The court interpreted the FHAA and ADA to prohibit indemnification by following the reasoning of *Northwest Airlines* as adopted in *Quality Built*.¹⁸¹

Most importantly, the *Archstone* court expressly denounced state breach of contract claims that Archstone asserted as a result of Niles Bolton’s failure to uphold its bargain and indemnify Archstone.¹⁸² The court reasoned that federal law preempts state law express contractual indemnification claims because such claims conflict with the FHAA and ADA.¹⁸³ More specifically, the court held that “Archstone’s express indemnity claim, based on its contract with Niles Bolton, is barred by federal law, every bit as much as its implied indemnity claim is barred.”¹⁸⁴ The court further reasoned that

173. *Id.*

174. *Id.*

175. *Id.* at 817.

176. *Id.*

177. Archstone’s predecessor actually contracted with Niles Bolton. Archstone acquired ownership in the apartment complexes through a buyout. *Id.*

178. *Id.*

179. See *Archstone-Smith Trust and Archstone-Smith Operating Trust’s Cross-Claim Against Niles Bolton Associates, Equal Rights Ctr. v. Archstone Smith Trust*, 603 F. Supp. 2d 814 (2009) (No. 04 Civ. 3975).

180. *Archstone*, 603 F. Supp. 2d at 824-26.

181. *Id.*

182. *Id.*

183. *Id.*

184. *Id.* at 825.

[a]s a matter of law, Archstone's state law claims for breach of contract and professional negligence are wholly derivative of Archstone's primary liability and are therefore what federal law regards as *de facto* claims for indemnification. Accordingly, those state law claims are barred because any recovery by Archstone would frustrate the achievement of Congress'[s] purposes in the FHA and the ADA.¹⁸⁵

Although none of the courts addressing indemnification in the FHAA and ADA context have clearly expressed policy rationales for disallowing indemnification, lurking in the background is the idea that imposing a non-delegable duty of compliance will result in more buildings meeting FHAA and ADA mandates.¹⁸⁶ Theoretically, all parties involved in constructing a building will then have an incentive to be familiar with the FHAA and ADA requirements and will ensure that all standards are met.¹⁸⁷ As discussed below, there are a number of problems with this theory as it relates to express indemnification.

III. THE FHAA AND ADA SHOULD BE INTERPRETED TO ALLOW EXPRESS INDEMNIFICATION

Several factors affect statutory interpretation, but in general, courts consider Congress's intent as evidenced through statutory language and legislative history, the purpose of the legislation, general policy considerations, and prudential concerns.¹⁸⁸ The *Quality Built* court and the courts following its lead have looked to these factors in interpreting the FHAA and ADA to prohibit indemnification,¹⁸⁹ but the result can be scrutinized through analysis of the same factors.

A. Congress's Intent

1. *Statutory Language.*—Both the FHAA and the ADA lack specific language providing for or prohibiting indemnification.¹⁹⁰ Lack of specific language creates an ambiguity, leaving open the possibility that Congress did not

185. *Id.* at 824 (emphasis added).

186. See Berg & Hecker, *supra* note 12, at 7 (citing *Paralyzed Veterans of Am. v. Ellerbe Becket Architects & Eng'rs, P.C.*, 945 F. Supp. 1, 2 (D.D.C. 1996) (supporting potential liability for all major players in design and construction)).

187. See *id.*

188. See DEBORAH B. MCGREGOR & CYNTHIA M. ADAMS, THE INTERNATIONAL LAWYER'S GUIDE TO LEGAL ANALYSIS AND COMMUNICATION IN THE UNITED STATES 193-206 (2008).

189. See, e.g., *United States v. Quality Built Constr., Inc.*, 309 F. Supp. 2d 767, 778-80 (E.D.N.C. 2003) (considering statutory language and structure, congressional intent, and legislative history); *Archstone*, 603 F. Supp. 2d at 822 (focusing on Congress's intent, which may be inferred from "the language of the statute itself, the statute's legislative history, the purpose and structure of the statute, and the likelihood that Congress intended to supersede or to supplement existing state remedies").

190. *Archstone*, 603 F. Supp. 2d at 822 (finding that "no such expressed right exists under [the FHAA and ADA]").

intend to interfere with contractual indemnification.¹⁹¹ As discussed above, many states have enacted anti-indemnification statutes that limit the ability of parties to expressly indemnify against certain losses.¹⁹² Because express indemnification is generally an acceptable means of shifting risk, anti-indemnification statutes threaten to drastically change established contractual relationships.¹⁹³ Recognizing the potential threat to existing risk allocation, many state anti-indemnification statutes provide that contractual indemnification provisions existing at the time of the statute's enactment remain enforceable.¹⁹⁴ This allowance assures parties the benefit of their bargains and gives them necessary notice to adjust mechanisms for allocating risk in the future.¹⁹⁵

Like states, Congress knows how to enact anti-indemnification legislation.¹⁹⁶ But Congress failed to even mention the concept in the Acts, let alone give parties notice of such a drastic change in traditionally accepted risk shifting. Theoretically, negotiating parties pay some premium for indemnification according to the amount of potential damages the agreement shifts.¹⁹⁷ Because the ADA and FHAA are silent with respect to contractual indemnification, prior to the recent decisions discussed above,¹⁹⁸ construction project participants had no notice or reason to think that premiums paid to shift risk would be held unenforceable. By construing the ADA and FHAA to deny contractual indemnification, courts are depriving premium-paying parties of the benefit of their bargains.¹⁹⁹ Premium-receiving parties gain windfalls to the disadvantage of premium-paying parties. If Congress intended to prohibit contractual indemnification under the ADA and FHAA, it should have put parties on notice by including specific language indicating prohibition. The lack of specific anti-

191. See MCGREGOR & ADAMS, *supra* note 190, at 194-95 (quoting Justice Stevens's analysis of how to interpret statutes in *City of Rancho Palos Verdes v. Abrams*, 544 U.S. 113, 129 (2005)).

192. See *supra* Part I.B.

193. Steven G.M. Stein & Shorge K. Sato, *Advanced Analysis of Contract Risk-Shifting Provisions: Is Indemnity Still Relevant?*, 27 CONSTR. LAW. 5, 5 (2007).

194. See Schoumacher, *supra* note 47, at § 1(a).

195. See *id.*

196. Congress has mentioned indemnification in a number of statutes and specifically limited or denied indemnification in several instances. See, e.g., Small Business Access to Surety Bonding Survey Act of 1992, 15 U.S.C. § 694b (2006) (limiting indemnification in the form of surety bond guarantees in certain instances); Oil Pollution Act of 1990, 33 U.S.C. § 2710 (respecting oil pollution, “[n]o indemnification . . . shall be effective to transfer liability imposed under this Act from a responsible party or from any person who may be liable for an incident under this Act to any other person”); Mercury Export Ban Act of 2008, 42 U.S.C. § 6939f(e)(B) (denying and limiting indemnification in some instances for persons delivering elemental mercury).

197. See POSNER, *supra* note 58, at 107.

198. See *supra* Part II and text accompanying note 19.

199. Of course, parties doing business in jurisdictions that have interpreted the Acts to prohibit indemnification should be on notice going forward, but this will not aid parties that made bargains long before the current trend and are now finding themselves subject to suit. Parties in undecided jurisdictions face an unpredictable bargaining climate.

indemnification language indicates that Congress did not intend to change existing contractual relationships.

2. *Legislative History*.—In considering whether Congress intended the FHAA and ADA to prohibit indemnification, courts have briefly mentioned legislative history of the Acts as a potentially indicative source of Congress's intent.²⁰⁰ The *Quality Built* court determined that there was nothing in the legislative history indicating that Congress intended to allow indemnification as a remedy.²⁰¹ On the other hand, one court in the Eastern District of Pennsylvania quoted language from House Reports on the FHAA as support for the proposition that Congress actually intended to prohibit contractual indemnification.²⁰² The House Report details a need for additional federal enforcement power.²⁰³ The court concluded that "Congress, in discussing the need for enhancing remedies to combat discrimination in housing, determined that enforcement should be bolstered by giving HUD new powers, not by permitting co-defendants to sue each other for contribution [and indemnification]."²⁰⁴ However, looking to the language from the House Report, Congress failed to even mention indemnification.²⁰⁵ Certainly Congress was concerned about the existing enforcement power of HUD, but the concern focused more on getting the potentially responsible parties to court rather than which specific party should foot the bill.²⁰⁶ Furthermore, the House Report indicates that Congress intended to strengthen private enforcement rights rather than supplant them with increased

200. See, e.g., *United States v. Quality Built Constr., Inc.*, 309 F. Supp. 2d 767, 779 (E.D.N.C. 2003) ("There is nothing in the legislative history of the FHAA which states or implies a right to contribution [or indemnification] on behalf of Defendants.").

201. *Id.*

202. *United States v. Gambone Bros. Dev. Co.*, No. 06-1386, 2008 WL 4410093, at *8 (E.D. Pa. Sept. 25, 2008). See *infra* text accompanying note 205 for the quoted language.

203. See H.R. REP. NO. 100-711, at 16-17 (1988), reprinted in 1988 U.S.C.C.A.N. 2173, 2177-78. Congress found that

[e]xisting law has been ineffective because it lacks an effective enforcement mechanism

Under existing law, although HUD investigates housing discrimination complaints, it can use only "informal methods of conference, conciliation, and persuasion" in an attempt to resolve them. HUD can do no more than this and lacks the power even to bring the parties to the conciliation table. HUD cannot sue violators to enforce the law, as in other civil rights laws.

. . . Since its passage, however, a consensus has developed that the Fair Housing Act has delivered short of its promise because of a gap in its enforcement mechanism.

The gap in enforcement is the lack of a forceful back-up mechanism which provides an incentive to bring the parties to the conciliation table with serious intent to resolve the dispute then and there.

Id.

204. *Gambone Bros.*, 2008 WL 4410093, at *8 (citing H.R. REP. NO. 100-711, at 16).

205. See *supra* text accompanying note 205 for what Congress specifically mentioned.

206. See *supra* text accompanying note 205.

governmental enforcement.²⁰⁷ The *Gambone* court, and others following its lead, misread Congress's statements. Reading the grant of increased enforcement power as a prohibition of contractual indemnification is a far stretch of the imagination.

Altogether, just as with the express language of the Acts, Congress simply did not mention indemnification in the legislative history. As argued above, the failure to specifically exclude contractual indemnification strengthens rather than weakens the argument that contractual indemnification clauses should be enforced.²⁰⁸ Indemnification clauses are widely used in the construction industry.²⁰⁹ Contracting parties need to be on notice that traditionally negotiated deal terms will not be enforced.²¹⁰ The *Archstone* court mentioned that one of the factors in determining Congress's intent is "the likelihood that Congress intended to supersede or supplement existing state remedies."²¹¹ Disregarding an indemnification clause leads to breach of contract, a cause of action for which state law provides a remedy.²¹² But the *Archstone* court did not give this factor serious consideration.²¹³ Instead, the court assumed that Congress intended to prohibit indemnification and searched for positive evidence to the contrary.²¹⁴ Such an approach is the reverse of the *Archstone* court's own standard. Rather than presuming that Congress intended to prohibit indemnification, the standard business practices and remedies of the construction industry should be presumed valid, and courts should require plaintiffs to present evidence to the contrary. The legislative history does not provide evidence that Congress intended to prohibit indemnification.

B. Purposes of the Acts

Because the statutory language and legislative history of the Acts are ambiguous regarding indemnification, courts have relied on the purposes of the Acts in support of prohibition.²¹⁵

207. See H.R. REP. NO. 100-711 ("Section 813 continues the private right of action under existing law, but eliminates certain restrictions on the exercise of that right. . . . [It does so in part by extending] the statute of limitations from 180 days to . . . [two] years."); see also Garcia v. Brockway, 526 F.3d 456, 475 (9th Cir. 2008) (Fisher, J., dissenting). For the importance of private causes of action, see *Trafficante v. Metro. Life Ins. Co.*, 409 U.S. 205, 211 (1972).

208. See *supra* Part III.A.1.

209. See *supra* Part I.B.

210. See *supra* Part III.A.

211. Equal Rights Ctr. v. Archstone Smith Trust, 603 F. Supp. 2d 814, 822 (D. Md. 2009).

212. See *supra* Part I.B.

213. The *Archstone* court mentioned supplementing state law remedies as a factor but then focused on the *Quality Built* and *Northwest Airlines* analysis. See *Archstone*, 603 F. Supp. 2d at 822.

214. *Id.*

215. See, e.g., *id.* ("The same imperative of Congressional purpose applies to attempts to 'contract around' the 'non-delegable' duties imposed by the FHA.").

1. *Distilling the Purposes.*—Respecting accessibility discrimination, Congress enacted the FHAA and ADA to serve similar purposes. FHAA language provides that “[it] is the policy of the United States to provide, within constitutional limitations, for fair housing throughout the United States.”²¹⁶ Professor Schwemm noted that “[t]he FHAA’s ban on handicap discrimination was intended to be ‘a clear pronouncement of a national commitment to end the unnecessary exclusion of persons with handicaps from the American mainstream’” and that “Congress believed that ‘[t]he right to be free from housing discrimination . . . [was] essential to the goal of independent living.’”²¹⁷

As for the ADA, Congress found that

the continuing existence of unfair and unnecessary discrimination and prejudice denies people with disabilities the opportunity to compete on an equal basis and to pursue those opportunities for which our free society is justifiably famous, and costs the United States billions of dollars in unnecessary expenses resulting from dependency and nonproductivity.²¹⁸

Further, “the [n]ation’s proper goals regarding individuals with disabilities are to assure equality of opportunity, full participation, independent living, and economic self-sufficiency for such individuals.”²¹⁹

One of the key purposes of the FHAA and the overriding goal of the ADA is therefore to curb discrimination against persons with disabilities. This Note is primarily concerned with curbing discrimination in the design and construction of buildings covered by the FHAA and ADA. All district courts that have considered whether indemnification should be foreclosed have implicitly or explicitly determined that discrimination in design and construction is best eliminated by making all parties to the project potentially liable regardless of fault or prior agreement.²²⁰ The *Archstone* court recited several of the ADA’s purposes as outlined above and stated that “[t]hese goals would be undermined if parties could simply ‘contract around’ their responsibilities under the statute.”²²¹ Similarly, the Southern District of Florida held that “[this] Court cannot approve an arrangement where a developer of a hotel can essentially contract around ADA compliance.”²²² The *Quality Built* court held that “[t]o allow Defendants to seek indemnity . . . would run counter to the purpose of the FHAA and undermine the regulatory goal by allowing the builder to escape any liability for violating the

216. 42 U.S.C. § 3601 (2006).

217. Schwemm, *supra* note 105, at 756 (citation omitted) (quoting H.R. REP. NO. 100-711, at 18 (1988), reprinted in 1988 U.S.C.C.A.N. 2173, 2179).

218. 42 U.S.C. § 12101(a)(8).

219. *Id.* § 12101(a)(7). For additional clarity on the purposes of the ADA, its explicit purposes are set out in 42 U.S.C. § 12101(b).

220. See *supra* text accompanying note 19.

221. Equal Rights Ctr. v. Archstone Smith Trust, 603 F. Supp. 2d 814, 825 (D. Md. 2009).

222. Access 4 All, Inc. v. Atl. Hotel Condos. Ass’n, Inc., No. 04-61740-Civ., 2005 WL 5643878, at *13 (S.D. Fla. Nov. 23, 2005).

Act.”²²³

2. *Two Issues Related to Achieving the Purposes.*—There are really two separate concerns related to curbing accessibility discrimination: (1) paying to bring noncompliant buildings that were subject to FHAA and ADA provisions during construction into compliance; and (2) preemptively ensuring that new construction projects are built according to FHAA and ADA mandates the first time around.²²⁴ Unfortunately, the courts do not explain why imposing liability on all parties to a project and prohibiting contractual indemnification would better serve the purposes of the Acts than imposing liability on all parties and allowing indemnification. The courts do not indicate which concern is best addressed by indemnification prohibition.

a. *Achieving compliance for current noncompliant structures.*—As for paying to bring noncompliant structures into compliance, the notion that parties can “contract around” the purposes of the Acts through indemnification is absurd. The mandates of the FHAA and ADA concern overall accessibility for persons with disabilities rather than who foots the bill for bringing structures into compliance.²²⁵ Ultimately, at least one party must be responsible for meeting FHAA and ADA mandates regardless of any contractual indemnification agreements.²²⁶ In order to “contract around” the FHAA and ADA, parties would have to contract with all potential future plaintiffs—including the federal government.

Regardless of the existence of an indemnification provision, all parties may still be sued and have judgments rendered against them.²²⁷ Those parties will have to satisfy the judgment to bring buildings into compliance regardless of whether they are ever successful in obtaining reimbursement through indemnification.²²⁸ The purpose of bringing structures into compliance is therefore fulfilled. Indemnified parties are simply reimbursed for their portion of the judgment because they paid a premium *ex ante* to provide an incentive for the other party to ensure compliance with the FHAA and ADA.²²⁹

b. *Preemptive compliance.*—Because prohibiting contractual indemnification fails to help bring noncompliant structures into compliance, contractual indemnification prohibition must be primarily targeting preemptive compliance. Apparently courts reason that by making all parties potentially liable for noncompliant structures without an option for reimbursement, all parties will be

223. United States v. Quality Built Constr., Inc., 309 F. Supp. 2d 767, 779 (E.D.N.C. 2003).

224. See, e.g., United States v. Pac. Nw. Elec., Inc., No. 01-019, 2003 U.S. Dist. LEXIS 7990, at *46 (D. Idaho Mar. 19, 2003) (“The relief sought by Plaintiff is aimed not only at preventing future violations of the FHA by Defendants, but also at removing the lingering effects of any past violations by requiring retrofitting necessary to bring the complexes into compliance.”).

225. See *supra* notes 218-22 and accompanying text.

226. Because indemnification is derivative, at least one party must be primarily liable. See *supra* Part I.B.

227. See *supra* Part I.B.

228. See *id.*

229. See *id.*

properly incentivized to ensure compliance with the Acts.²³⁰ But prohibiting indemnification is a radical change in the traditional construction industry structure that can produce the opposite consequence.²³¹ In less complex industries, providing such an all-encompassing incentive may prove most effective.²³² However, because of the heavily contract-based, diversified nature of the construction industry, incentives are potentially skewed to the detriment of the purposes of the Acts.

By creating a non-delegable duty for all parties involved in a project to comply with the Acts, designers have less incentive to ensure that building designs follow all regulations.²³³ If a designer has indemnified an owner against potential noncompliance because of faulty design or poor supervision of the construction process, there is no question that the designer has the utmost incentive to ensure compliant design and construction.²³⁴ Without risk shifting through indemnification, designers have less incentive to ensure compliance in design and through supervision because they may ultimately split the cost of any problematic design with all other parties.²³⁵ The same reasoning applies to builders who agree to indemnify owners.²³⁶ Without indemnification provisions, builders have less incentive to ensure compliance because of cost splitting.²³⁷

Owners are the most likely target of an anti-indemnification interpretation because, as they have limited expertise and resources for design and construction, they are the parties most likely to seek out indemnification from designers and builders.²³⁸ But prohibiting indemnification does not change the elemental nature of the construction industry.²³⁹ Most owners simply do not have the expertise to

230. See Berg & Hecker, *supra* note 12, at 5.

231. See Stein & Sato, *supra* note 195, at 5.

232. For an example of a less complex industry, see the discussion of discriminatory employment practices in Part II.B.

233. See FEINMAN, *supra* note 40, at 278 (explaining that when courts tamper with contractual obligations of construction project participants, “parties cannot accurately predict to whom and for what they will be liable, [and therefore] it is impossible for them to plan appropriately for performance and risk in the course of construction”). Of course, designers will continue to have some incentive to comply because their reputation depends on quality service, and failure to design compliant structures may lead to negligence suits. See FUNDAMENTALS OF CONSTRUCTION LAW, *supra* note 26, at 49-50, 54-55.

234. See FEINMAN, *supra* note 40, at 277-78.

235. See John E. Calfee & Richard Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards*, 70 VA. L. REV. 965, 978 (1984).

236. See *id.*

237. See *id.*

238. See CONSTRUCTION LAW, *supra* note 11, at 131-33.

239. It is possible that increasing costs to owners by prohibiting risk shifting could lead to more prominent developers bringing designers and builders in-house. They may do so if the costs of exercising enough control over the project to ensure compliance (if this is possible) exceed the costs of bringing designers and builders in-house. See *supra* note 36.

design and construct compliant buildings.²⁴⁰ They pay for entities with expertise to design, monitor, and construct compliant buildings.²⁴¹ They rely on added contractual incentives through indemnification to ensure performance.²⁴² Of course, if an owner blatantly instructs a designer or builder to disregard the mandates of the Acts, the owner should not be reimbursed for liability. In cases of such willful behavior, indemnification provisions will be void anyway.²⁴³ At best, prohibiting contractual indemnification results in owners paying more for monitoring—with no guarantee that the results will change—and passing off the excess costs to ultimate users of facilities.

By prohibiting indemnification, courts have diffused costs among all participants and created uncertainty rather than allowing parties to concentrate responsibility with those best able to avoid the cost. In particular, the parties that are in the best position to avoid costs are primarily the designers and, to a lesser extent, the builders.²⁴⁴ As in *Archstone*, parties considering the problem *ex ante* are likely to allocate the risk of noncompliance to the designer.²⁴⁵ Since the designer is in the best position to avoid liability under the FHAA and ADA in the first place, such an allocation of responsibility is most efficient.²⁴⁶ This arrangement will ultimately save all involved—including society as a whole—costs associated with bringing existing structures into compliance or building the structure from the outset being passed on to future users of the facilities.²⁴⁷

The most effective path to curbing discrimination in accessibility design and construction is to provide a proper incentive structure through increased enforcement of the Acts coupled with contractual indemnification. Relaxed enforcement policies give no parties an incentive to ensure compliance with the Acts.²⁴⁸ Stringent enforcement incentivizes the party with the most financial

240. See *supra* Part I.A.

241. See *id.*

242. See *supra* Part I.B.

243. See *id.*

244. See James P. Colgate, *If You Build It, Can They Sue? Architects' Liability Under Title III of the ADA*, 68 FORDHAM L. REV. 137, 160-63 (1999).

245. See Equal Rights Ctr. v. Archstone Smith Trust, 603 F. Supp. 2d 814, 817 (D. Md. 2009). Parties in the best position to avoid the risk of noncompliance should be eager to accept a premium because they have the ability to easily avoid payout. See *infra* text accompanying note 249.

246. For a discussion of how efficiency is best achieved by allocating responsibility to parties in the best position to avoid costs, see R.H. Coase, *The Problem of Social Cost*, 3 J.L. & ECON. 1, 42-44 (1960). See also POSNER, *supra* note 58, at 106 (determining that a hypothetical contractor is in a better position to prevent fire in the construction of a building than the owner because “he is in a better position . . . to estimate the likelihood and consequences of fire at various stages in the construction” and “he controls the premises and is knowledgeable about the fire hazards of buildings under construction”).

247. See Colgate, *supra* note 244, at 160-63.

248. See Calfee & Craswell, *supra* note 235, at 976-82.

resources at stake if liability is imposed.²⁴⁹ Allowing contractual indemnification creates an incentive within the construction relationship for parties with the greatest expertise and control to ensure compliance.²⁵⁰ Stringent external enforcement of FHAA and ADA mandates accentuates the incentive on the parties that agreed to accept responsibility through contractual indemnification.

A simple hypothetical illustrates the proper incentive structure. A store owner leaves the cashier (*C*), the salesman (*S*), and the janitor (*J*) in charge of the store for several days at a time. In the event that money is missing from the cash register, the owner requires all workers to repay the deficit. *C*, *S*, and *J* agree that because *C* is in the best position to watch the cash register, if any money is missing, all three will contribute funds to restore the deficit, but *C* must repay *S* and *J*. If the owner never makes any worker pay even when money is missing, *C*, *S*, and *J* really do not care who bears ultimate responsibility. If the owner always makes all workers contribute but fails to make *C* reimburse *S* and *J*, *C* is more likely to be watchful but can afford to shirk because *S* and *J* are shouldering a portion of his financial burden. If the owner always makes all workers contribute and enforces the private agreement, *C* would be a fool to shirk.

In a typically complicated construction project with diversified responsibility and expertise, placing incentives on parties with little knowledge and practical control over accessibility design does not help accomplish the purposes of the Acts. On the whole, parties to the construction project are not “contracting around” the FHAA and ADA specifications because at least one of the parties must remain liable for properly following the guidelines.²⁵¹ This represents a shift of responsibility from one party to another, making the process more efficient and more likely to be adequately performed.²⁵² It does not allow parties to disregard the mandates of the FHAA and ADA.

C. Public Policy and Other Considerations

1. Public Policy.—A number of articles have criticized any potential escape from liability by designers for FHAA and ADA structures as bad public policy.²⁵³ The underlying rationale is that designers are the most influential and knowledgeable parties to design and supervise the construction of FHAA- and ADA-compliant facilities.²⁵⁴ Designers are plainly in the best position to avoid

249. *See id.*

250. *See supra* Part I.B.

251. *See id.*

252. *See* POSNER, *supra* note 58, at 106.

253. *See, e.g.*, Colgate, *supra* note 244, at 160-63.

254. *Id.* at 161.

The responsibility for proper building design is thus ascribed not to clients, who may have limited knowledge of design regulations and little incentive to meet them, but to licensed architects, whose “training and professional status place them in the best position to protect the public by assuring that their designs safeguard life, health, and property to the fullest extent possible.”

the cost of noncompliance.²⁵⁵ If the design is noncompliant at the planning stage, it costs comparatively little to fix as opposed to, for instance, near construction completion when the builder notices a noncompliant aspect of the design.²⁵⁶ Designers also play a supervisory role in the construction process to ensure that plans are properly implemented.²⁵⁷ Thus, not only do designers make the initial planning decisions, but they are also in a position to monitor the builder and catch negligent errors.²⁵⁸

Noncompliant structures waste time and resources that could be better used to serve society's needs.²⁵⁹ Of course, one of the overriding purposes of the Acts, curbing discrimination against persons with disabilities, is itself a strong public policy.²⁶⁰ Contractual indemnification adds to the argument in favor of greater responsibility for designers by properly incentivizing designers to ensure complete compliance.²⁶¹ Prohibiting indemnification under the ADA and FHAA cracks a door for designers to partially escape liability. As a result, structures are more likely to be noncompliant, costing society valuable time and resources.

2. *Role of Insurance.*—The construction industry relies heavily on insurance for all aspects of operations.²⁶² Because construction insurance is a complicated field,²⁶³ this Note is not meant to provide a thorough analysis of insurance coverage options. It is enough to simply highlight the impact insurance has on the problem of liability for noncompliant structures under the FHAA and ADA.

Neither the FHAA nor the ADA explicitly restricts construction project participants from obtaining an insurance policy to cover potential liability for

Id. (quoting William David Flatt, Note, *The Expanding Liability of Design Professionals*, 20 MEM. ST. U. L. REV. 611, 615 (1990)).

255. See *LEGAL HANDBOOK FOR ARCHITECTS, ENGINEERS AND CONTRACTORS: ISSUES IN CONSTRUCTION AND TECHNOLOGY 2002/2003*, at 213-18 (Albert H. Dib ed., 2003).

256. The technology and industrial engineering fields analyze such decisions using cost of change curves. These curves describe an exponentially increasing cost as the development process continues. See James E. Folkestad & Russell L. Johnson, *Resolving the Conflict Between Design and Manufacturing: Integrated Rapid Prototyping and Rapid Tooling (IRPRT)*, 17 J. INDUS. TECH. 1, 3-4 (2001). Though the curve for building designers may be steeper or flatter, there should be a reasonably similar relationship.

257. See *FUNDAMENTALS OF CONSTRUCTION LAW*, *supra* note 26, at 8 ("A less well-known, but related and critical expectation of the owner is that the architect will participate in the administration of the construction contract to assure that the design objectives of the owner are fulfilled.").

258. *See id.*

259. Colgate, *supra* note 244, at 162.

260. *See supra* Part III.B.1.

261. *See supra* Part III.B.2.b.

262. See *CONSTRUCTION LAW*, *supra* note 11, at 531-42.

263. SMITH, CURRIE & HANCOCK'S COMMON SENSE CONSTRUCTION LAW: A PRACTICAL GUIDE FOR THE CONSTRUCTION PROFESSIONAL 440 (Thomas J. Kelleher, Jr. ed., 3d ed. 2005) ("Insurance planning for construction projects is extremely complex and specialized . . .").

noncompliance.²⁶⁴ The FHAA is silent regarding insurance coverage.²⁶⁵ The ADA specifically states:

Subchapters I through III of this chapter and title IV of this Act shall not be construed to prohibit or restrict—

(1) an insurer, hospital or medical service company, health maintenance organization, or any agent, or entity that administers benefit plans, or similar organizations from underwriting risks, classifying risks, or administering such risks that are based on or not inconsistent with State law . . .²⁶⁶

Accordingly, construction project parties are free to bargain for insurance provisions that cover losses associated with ADA and FHAA noncompliance.

*Washington Sports and Entertainment v. United Coastal Insurance Co.*²⁶⁷ illustrates potential shifting of liability through insurance policies.²⁶⁸ United Coastal Insurance (UCI) issued a design professional insurance policy to the designer of a sports complex.²⁶⁹ The sports complex owner paid for the policy and was named as an additional insured.²⁷⁰ Originally, the policy excluded “any and all claims both from victims and governmental agencies arising out of or relating to . . . violation of the Americans with Disabilities Act or violation of any other civil rights.”²⁷¹ But the designer and owner paid a substantial premium²⁷² to amend the policy so that the exclusion did not “apply to a design error that could result in a violation of the Americans with Disabilities Act.”²⁷³ After project completion, the Paralyzed Veterans of America successfully brought suit alleging noncompliant design.²⁷⁴ Despite its contractual promise, UCI failed to defend the suit or indemnify the owner for costs to bring the facility into compliance.²⁷⁵ The owner sued UCI, and the court held that the owner “paid a substantial premium to shift risks to . . . [UCI, and UCI] willingly agreed to accept those risks. Without showing that [the owner’s and designer’s] motives stepped beyond a general risk aversion, [UCI] . . . cannot evade its duty to honor the [p]olicy and to defend plaintiffs . . .”²⁷⁶

264. For the lack of specific language in the FHAA concerning insurance coverage for failures to design and construct compliant buildings, see 42 U.S.C. §§ 3601-3619 (2006). For the ADA’s stance on insurance coverage, see 42 U.S.C. § 12201(c)(1) and *infra* note 266.

265. 42 U.S.C. §§ 3601-3619.

266. *Id.* § 12201(c)(1).

267. 7 F. Supp. 2d 1 (D.D.C. 1998).

268. *See id.* at 11.

269. *Id.* at 3-4.

270. *Id.* at 3.

271. *Id.* at 5.

272. The owner paid over \$500,000 up front for the policy. *Id.* at 4.

273. *Id.* at 5.

274. *Id.* at 3.

275. *Id.*

276. *Id.* at 11.

Because insurance policies and contractual indemnification clauses are both forms of contractual indemnification that shift risk,²⁷⁷ prohibiting one while enforcing the other is absurd. On the surface, there is a difference between insurance policies and contractual indemnification clauses—the source of ultimate payment for noncompliant structures.²⁷⁸ If an insurance company honors a claim, the cost is spread to all other premium-paying clients of the company.²⁷⁹ If the insurance company is large, the bill may be paid by a large segment of society.²⁸⁰ If the parties contractually shift risk through indemnification clauses, the party assuming the risk pays the ultimate bill.²⁸¹ But in reality, a wise risk-assuming party would require a premium to be paid in the contract and in turn take that premium and buy an insurance policy to cover its potential liability. Depending on the availability and cost of insurance coverage, parties can achieve similar results as those arrived at through contractual indemnification.

United Coastal was decided in 1998,²⁸² five years before the *Quality Built* decision was handed down and ten years before the recent flurry of cases following *Quality Built*'s lead.²⁸³ Courts may soon face the same issue as found in *United Coastal* but with an argument that express indemnification is prohibited under the Acts. The argument follows that in order to be consistent, courts prohibiting contractual indemnification under the Acts would likewise have to prohibit parties from obtaining insurance coverage. But neither the FHAA nor the ADA explicitly prohibits parties from obtaining insurance coverage in such situations.²⁸⁴ It would be illogical for Congress to have allowed insurance companies to indemnify against FHAA and ADA liability while simultaneously disallowing contractual indemnification between construction project parties. Congress did not intend to create an inconsistent system. By prohibiting contractual indemnification, courts have fashioned the inconsistency.

3. *Contractual Indemnification Prohibition Is at Odds with Policy of Some States.*—Accessibility law is by no means exclusively a federal affair.²⁸⁵ Most states have some form of accessibility standards, many of which are more rigorous than the FHAA and ADA mandates.²⁸⁶ The FHAA provides that

277. See *id.* (discussing insurance as a tool for shifting risk).

278. See ROBERT H. JERRY II & DOUGLAS R. RICHMOND, UNDERSTANDING INSURANCE LAW 13-14 (4th ed. 2007).

279. See *id.*

280. See *id.*

281. See *supra* Part I.B.

282. Wash. Sports & Ent'mt v. United Coastal Ins. Co., 7 F. Supp. 2d, 1 (D.D.C. 1998).

283. *Quality Built* was decided in 2003. United States v. Quality Built Constr., Inc., 309 F. Supp. 2d 767, 778-79 (E.D.N.C. 2003). For the dates of decisions handed down since *Quality Built*, see *supra* text accompanying note 19.

284. See *supra* text accompanying note 264 and accompanying text.

285. See Schwemm, *supra* note 105, at 762.

286. See *id.* at 809; see generally A STATE-BY-STATE GUIDE TO CONSTRUCTION & DESIGN LAW: CURRENT STATUTES AND PRACTICES (Carl J. Circo & Christopher H. Little eds., 1998) (comparing accessibility laws, among others, across the United States).

“[n]othing in this subchapter shall be construed to invalidate or limit any law of a State . . . that requires dwellings to be designed and constructed in a manner that affords handicapped persons greater access than is required by this subchapter.”²⁸⁷ The ADA contains a very similar provision.²⁸⁸ Accordingly, Congress showed some deference to states with requirements equal to or more stringent than the federal mandates.²⁸⁹ Yet some states that have mirror-image or more aggressive accessibility guidelines do not prohibit contractual indemnification between construction project parties.²⁹⁰ A glimpse at Indiana law illustrates the point.

To date, Indiana has not weighed in on whether the FHAA or ADA prohibits contractual indemnification between construction project parties.²⁹¹ At the state level, Indiana’s legislature essentially adopted a mirror image of the FHAA as a state version of anti-housing discrimination legislation.²⁹² Indiana also incorporated the ADA guidelines for building standards.²⁹³ In addition, Indiana requires that designers certify design compliance with all building codes in order for owners to obtain state-issued design releases prior to beginning a construction project.²⁹⁴ Requiring design releases places an increased burden on designers of construction projects, rather than builders and owners, to comply with applicable standards by requiring submission of plans and promise of compliance.²⁹⁵ By inference, Indiana therefore recognizes that building designers have more specialized skills and control over the project’s accessibility features than builders and owners.

Despite mirror-image standards of the FHAA and ADA and increased architect responsibility, Indiana does not prohibit contractual indemnification for

287. 42 U.S.C. § 3604(f)(8) (2006); *see also id.* § 3615, providing:

Nothing in this subchapter shall be construed to invalidate or limit any law of a State or political subdivision of a State, or of any other jurisdiction in which this subchapter shall be effective, that grants, guarantees, or protects the same rights as are granted by this subchapter; but any law of a State, a political subdivision, or other such jurisdiction that purports to require or permit any action that would be a discriminatory housing practice under this subchapter shall to that extent be invalid.

288. *Id.* § 12201(b). This section provides:

Nothing in this chapter shall be construed to invalidate or limit the remedies, rights, and procedures of any Federal law or law of any State or political subdivision of any State or jurisdiction that provides greater or equal protection for the rights of individuals with disabilities than are afforded by this chapter.

289. *See id.*

290. *See, e.g., IND. CODE §§ 22-9.5-1-1 to 22.9.5-11-3 (2010); id. § 26-2-5-1.*

291. The author is unaware of any Indiana cases interpreting the FHAA or ADA regarding indemnification.

292. *See IND. CODE §§ 22-9.5-1-1 to 22-9.5-11-3.* In particular, Section 22-9.5-1-1(3) provides that one purpose of Indiana’s version of fair housing legislation is “[t]o provide rights and remedies substantially equivalent to those granted under federal law.”

293. *See id.* § 22-13-4-1.5.

294. *See id.* § 22-15-3-1-2.

295. *See id.* § 22-15-3-2.

noncompliant design and construction. Unlike the FHAA and ADA, Indiana statutes specifically address the enforceability of indemnification agreements between owners, builders, and designers and do not prohibit all indemnification agreements for failure to comply with Indiana's version of the FHAA and ADA.²⁹⁶ Instead, Indiana clearly signaled which agreements are and are not enforceable by prohibiting indemnification resulting "from the sole negligence or willful misconduct of the promisee or promisee's agents, servants or independent contractors."²⁹⁷ Accordingly, owners, builders, and designers may allocate responsibility for accessibility compliance between the parties to the project as long as noncompliance does not result from the sole negligence or willful misconduct of the party seeking indemnification.²⁹⁸

As an illustration of the potential conflict between state and federal express indemnification rules, litigation based on the same noncompliant feature in an Indiana building could result in disparate treatment depending on whether the claim was brought under the federal Acts or state law equivalents. In order for Indiana to maintain consistent treatment of contractual relationships, Indiana federal courts would have to interpret the FHAA and ADA to allow contractual indemnification. Otherwise, if Indiana federal courts follow *Archstone*'s lead, the state law claims are preempted by federal law even though the parties may have been relying on state law mandates.²⁹⁹ Even if the parties bring a state law breach of contract action rather than an indemnification action, to the extent that the contract action is based on liability under the FHAA and ADA, the contract is unenforceable.³⁰⁰ Congress gave no express consideration of such a direct conflict.³⁰¹

D. Prudential Considerations

One court that interpreted the Acts to prohibit indemnification³⁰² cited the Seventh Circuit Court of Appeals for the proposition that absent Congress expressly giving a right of contribution or indemnification, none should be afforded by the court.³⁰³ The Seventh Circuit's reasoning for this approach was based on prudential considerations.³⁰⁴ The court recognized that cases involving cross-claims for contribution and indemnification create more work for the court and complicate the proceeding.³⁰⁵

296. See *id.* § 26-2-5-1.

297. *Id.*

298. See *id.*

299. Equal Rights Ctr. v. Archstone Smith Trust, 603 F. Supp. 2d 814, 825 (D. Md. 2009).

300. *Id.*

301. See 42 U.S.C. §§ 3601-3619 (2006) (FHAA language); *id.* § 12101-12213 (ADA language).

302. United States v. Shanrie Co., 610 F. Supp. 2d 958, 960 (S.D. Ill. 2009).

303. Anderson v. Griffin, 397 F.3d 515, 523 (7th Cir. 2005).

304. *Id.*

305. *Id.*

While the Seventh Circuit's prediction of increased courtroom complication is likely correct in the case of allowing implied indemnification and contribution claims,³⁰⁶ express indemnification is more likely to lead to the opposite result.³⁰⁷ Perhaps the greatest virtue of contractual indemnification is predictability.³⁰⁸ Uncertainty in the distribution of damages gives parties an incentive to fight things out in court because they may have something to gain at the expense of other parties.³⁰⁹ When parties come to an enforceable agreement respecting damage allocation ex ante, uncertainty is greatly reduced—and likewise, so is the probability of payoff through lawsuits.³¹⁰ Parties facing relative certainty in damage allocation are more likely to settle their dispute outside of court or in the early stages of litigation.³¹¹ Needless litigation is therefore avoided, and society's scarce resources can be put to better use.

CONCLUSION

There is a disconnect between attempting to curb accessibility discrimination and interpreting the FHAA and ADA to prohibit contractual indemnification. The increasing number of cases and commentary on the subject indicates an elevated concern for enforcement of accessibility legislation in the design and construction of applicable facilities.³¹² Improving accessibility so that persons with disabilities can enjoy a richer life is of utmost importance. Heightened awareness of the problem and enforcement of the Acts is long overdue.

Improved accessibility is best achieved by providing the proper incentive structure. The proper incentive structure combines increased overall enforcement of design and construct mandates with internal risk allocation through contractual indemnification. With increased overall enforcement, the party that bears ultimate financial responsibility for the cost of noncompliance is more likely to ensure proper performance.³¹³ Through contractual indemnification, construction project participants can place ultimate financial responsibility on parties in the best position to avoid costs.³¹⁴ Increasing enforcement but prohibiting indemnification does more harm than good because even though all parties are

306. Courtroom complexity increases with implied indemnification and contribution claims because they are essentially imposing another trial within a trial. The court must make determinations ex post about the relative fault of the parties and how much damage, if any, should be shifted. *See id.*

307. *See Robert Cooter, Stephen Marks, & Robert Mnookin, Bargaining in the Shadow of the Law: A Testable Model of Strategic Behavior*, 11 J. LEGAL STUD. 225, 237-39 (1982).

308. *See supra* Part I.B.

309. *See generally* Cooter et al., *supra* note 307.

310. *See id.*

311. *See id.*

312. Schwemm, *supra* note 105, at 754-56.

313. *See Calfee & Craswell, supra* note 235, at 976-82.

314. *See supra* Parts I.A. & I.B.

more fearful of liability, they know that costs may be split.³¹⁵ Congress did intend to curb accessibility discrimination, but it did not intend to skew the internal incentive structure of the construction industry.³¹⁶ By allowing parties to contractually allocate risk and responsibility through indemnification and simultaneously ratcheting up enforcement, persons with disabilities will benefit, society will benefit, and the purposes of the FHAA and ADA will be better served.

315. See Calfee & Craswell, *supra* note 235, at 976-82.

316. See *supra* Parts III.A. & III.B.

APPLICATION OR REGISTRATION?: CONFUSION REGARDING THE COPYRIGHT ACT'S PREREQUISITE TO COPYRIGHT INFRINGEMENT LAWSUITS

THOMAS M. LANDRIGAN*

INTRODUCTION

Section 411(a) of the Copyright Act of 1976¹ (“the Act”) mandates that “no civil action for infringement of the copyright in any United States work shall be instituted until preregistration or registration of the copyright claim has been made in accordance with this title.”² Courts interpreting the 411(a) requirement are split between two opposing approaches to interpreting this provision.³ Proponents of the “registration approach” read this section literally, requiring a potential plaintiff to wait until the United States Copyright Office (“Copyright Office”) actually registers (or denies registration of) the copyright.⁴ Proponents of the “application approach,” on the other hand, believe that for reasons of policy and practicality, submission of an application for registration is enough to allow the plaintiff to bring a claim.⁵

Not only are federal courts split over the meaning of the registration requirement, but cases dealing with the application versus registration issue evince a notable lack of clarity among federal judges regarding the issue and the legal precedent to follow.⁶ As one commentator remarked, “Confused as to what events must normally occur in order for copyright registration to be achieved under federal law? So are the federal courts.”⁷ This Note focuses specifically on cases in the Seventh and Eighth Circuits in which district courts and commentators have struggled to understand the meaning of unclear circuit court opinions.⁸

In *Reed Elsevier, Inc. v. Muchnick*,⁹ the Supreme Court recently addressed the question of whether section 411(a) is a jurisdictional prerequisite or whether federal courts have discretion to hear cases when 411(a) has not been satisfied, such as the class action settlement approval which the parties in that case

* J.D. Candidate, 2011, Indiana University School of Law—Indianapolis; B.A., 2008, Purdue University, West Lafayette, Indiana.

1. 17 U.S.C. §§ 101-810 (2006 & Supp. 2009).

2. *Id.* § 411(a) (Supp. 2009).

3. *La Resolana Architects, PA v. Clay Realtors Angel Fire*, 416 F.3d 1195, 1201-02 (10th Cir. 2005), abrogated by *Reed Elsevier, Inc. v. Muchnick*, 130 S. Ct. 1237 (2010).

4. *Id.* at 1202.

5. *Id.* at 1203.

6. See discussion *infra* Part III.

7. Mose Bracey, Note, *Searching for Substance in the Midst of Formality: Copyright Registration as a Condition Precedent to the Exercise of Subject-Matter Jurisdiction by Federal Courts over Copyright Infringement Claims*, 14 J. INTELL. PROP. L. 111, 114 (2006).

8. See discussion *infra* Part III.

9. 130 S. Ct. 1237 (2010).

sought.¹⁰ The Second Circuit had previously held that 411(a) is a jurisdictional requirement and that each plaintiff's claim would have to satisfy this requirement for a federal court to have jurisdiction to approve the settlement, noting "widespread agreement among the circuits that section 411(a) is jurisdictional."¹¹ The Supreme Court reversed, holding that 411(a) is a "precondition to filing a claim that does not restrict a federal court's subject-matter jurisdiction."¹² Thus, the Court decided that district courts may adjudicate disputes concerning unregistered works, and district court approval of this class action settlement was proper although the class included authors who had not registered.¹³ Yet the Court declined to address the application versus registration issue—the question of whether filing an application satisfies this "precondition" to filing an infringement suit, or whether section 411(a) requires a decision by the Copyright Office.¹⁴

This Note examines judicial interpretation of the Copyright Act's registration requirement. Part I provides background information about copyright law and the mechanics and purpose of registration. It also discusses the recently added option of preregistration for certain types of works. Part II discusses the disagreement between federal courts' interpretations of the registration requirement as well as the various rationales underlying the two approaches. It also critiques each of the names for the "registration" approach,¹⁵ alternatively referred to as the "issuance" or "approval" approach.¹⁶ Part III analyzes the causes for confusion in the Seventh and Eighth Circuits. Finally, Part IV suggests that courts should follow the registration approach as followed by the Tenth Circuit as long as section 411(a) of the Copyright Act remains in its current form. It goes on to recommend, however, that Congress alter the registration requirement in section 411(a) so that this precondition is satisfied as

10. *Id.* at 1243.

11. *In re Literary Works in Elec. Databases Copyright Litig.*, 509 F.3d 116, 121-22 (2d Cir. 2007), *rev'd sub nom.*, *Reed Elsevier, Inc. v. Muchnick*, 130 S. Ct. 1237 (2010).

12. *Reed Elsevier*, 130 S. Ct. at 1241.

13. *Id.* at 1242, 1249.

14. Cases discussed in this Note refer to the 411(a) requirement as jurisdictional because they precede the Supreme Court's *Reed Elsevier* decision. This Note does not attempt to redact all such "jurisdictional" references, but the reader should keep in mind that in light of *Reed Elsevier*, the 411(a) requirement is no longer characterized as restricting the district courts' subject-matter jurisdiction.

15. It is important to keep in mind the distinction between the term "registration" as required under section 411(a) of the Act and "registration" that means the Copyright Office has reviewed and actually registers the application. As discussed throughout this Note, the registration approach sees these meanings as the same, while the application approach interprets the statutory use of "registration" loosely so that application suffices without actual registration by the Copyright Office. *See discussion infra* Part II.

16. *See, e.g.*, *Bracey, supra* note 7, at 127 ("issuance approach"); *Erin Hogan, Approval Versus Application: How to Interpret the Registration Requirement Under the Copyright Act of 1976*, 83 DENVER U. L. REV. 843, 854-55 (2006) ("approval approach").

soon as a copyright holder has submitted his application materials for registration.

I. BACKGROUND OF RELEVANT COPYRIGHT LAW

Before the passage of the Copyright Act, unregistered copyrights were not federally recognized and were protected only to the extent provided by state property law.¹⁷ Following passage of the Act in 1976 and later the Berne Convention Implementation Act of 1988,¹⁸ federal law now recognizes copyright ownership without the requirement of registration, thus preempting state law.¹⁹ The Act removed previous formalities to copyright existence and established the creator's automatic copyright ownership in his work.²⁰ The Act states that "registration is not a condition of copyright protection."²¹ In other words, because federal law recognizes copyright ownership even before the creator takes any action to register, an unregistered copyright can be infringed but not sued upon. According to the Supreme Court, to show copyright infringement, a plaintiff must establish only that he owns a valid copyright and that the defendant has copied original elements in the work.²²

A. *The Purpose of Registration*

Although the law recognizes unregistered copyright ownership, an owner may realize certain benefits only by registering the copyright. The Act mandates that copyright owners register as a condition precedent to filing suit in federal court to protect their copyrights—"no civil action for infringement of the copyright in any United States work shall be instituted until preregistration or registration of the copyright claim has been made in accordance with this title."²³ Thus, a copyright can be valid, owned, and infringed,²⁴ yet its owner is not able to sue to enforce his copyright if he has not registered.²⁵ The Eighth Circuit summed up the primary purpose of registration by explaining that although "registration is required under section 411 of the Copyright Act in order to bring a suit for infringement, infringement itself is not conditioned upon registration of the copyright. Thus, a copyright holder can register a copyright and file suit after infringement occurs."²⁶ In addition to serving as a prerequisite to

17. Rita Marie Cain, *Timing Is Everything: Copyright Registration and Preregistration*, 88 J. PAT. & TRADEMARK OFF. SOC'Y 381, 381 (2006).

18. Pub. L. No. 100-568, 102 Stat. 2853 (1988) (codified as amended in scattered sections of 17 U.S.C. § 101).

19. Cain, *supra* note 17, at 381-82.

20. 17 U.S.C. § 102(a) (2006).

21. *Id.* § 408(a).

22. Feist Publ'n's, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991).

23. 17 U.S.C. § 411(a) (2006 & Supp. 2009).

24. See *id.* § 408(a) ("registration is not a condition of copyright protection.").

25. See *id.* § 411(a).

26. Olan Mills, Inc. v. Linn Photo Co., 23 F.3d 1345, 1349 (8th Cir. 1994) (internal citation

infringement suits, registration is also required in order to recover statutory damages and attorneys' fees under section 412.²⁷ According to the Eighth Circuit, "[t]he timing of registration only determines whether the copyright holder can recover statutory as opposed to actual damages."²⁸

To determine when a copyright owner is eligible for the benefits of registration, section 411(a) begs the question of when registration is made "in accordance with this title."²⁹ This question has created confusion among federal courts, which are split as to when registration has taken place.³⁰ The Act's circular definition of "registration" gives little guidance; "[r]egistration", for purposes of . . . [section 411, among others], means a registration of a claim in the original or the renewed and extended term of copyright."³¹ Oddly enough, registration does not have to be granted by the Copyright Office for the owner to enjoy the benefits of registration—section 411(a) also allows a copyright owner to institute suit where "the deposit, application, and fee required for registration have been delivered to the Copyright Office in proper form *and registration has been refused*" as long as the Register of Copyrights is given notice of the action and a copy of the complaint.³²

B. Registration Under the Copyright Act

The Act specifies the requirements for applying for registration: the copyright owner must deliver an application, a fee, and a deposit to the Copyright Office.³³ The deposit essentially consists of one or more copies of the work to be registered.³⁴ After a processing time that ranges up to around twenty months, the Copyright Office reviews the application materials and sends the applicant either a certificate of registration or a rejection notice.³⁵ According to the Act, "[t]he effective date of a copyright registration is the day on which an application, deposit, and fee, which are later determined by the Register of

omitted).

27. 17 U.S.C. § 412.

28. *Olan Mills*, 23 F.3d at 1349.

29. See 17 U.S.C. § 411(a) (Supp. 2009) ("no civil action for infringement . . . shall be instituted until preregistration or registration . . . has been made *in accordance with this title*.")(emphasis added).

30. See discussion *infra* Part II.

31. 17 U.S.C. § 101 (2006).

32. *Id.* § 411(a) (emphasis added).

33. *Id.* § 408(a).

34. *Id.* § 408(b).

35. *I've Submitted My Application, Fee, and Copy of My Work to the Copyright Office. Now What?*, U.S. COPYRIGHT OFFICE, <http://www.copyright.gov/help/faq/faq-what.html#certificate> (last modified Apr. 4, 2011) ("Most online filers should receive a certificate within 3.3 months. Many will receive their certificates earlier," and "[m]ost of those who file on these [paper] forms should receive a certificate within 11.5 months of submission. Many will receive their certificates earlier.").

Copyrights or by a court of competent jurisdiction to be acceptable for registration, have all been received in the Copyright Office.”³⁶ However, this provision is poorly drafted, and it is unclear whether the copyright registration date is effective at the moment the Copyright Office receives these items, or only retroactively effective after a registration decision has been made.³⁷

C. Preregistration

The text of section 411(a) actually specifies “preregistration or registration” as the prerequisite to copyright infringement suits.³⁸ The language “preregistration or” was added to the statute in 2005 as part of the Family Entertainment and Copyright Act of 2005 (FECA).³⁹ FECA essentially preempts the registration versus application approach issue for certain works, allowing suit for infringement of such works once an application for preregistration has been filed.⁴⁰ Preregistration is available for motion pictures, sound recordings, musical compositions, literary works being prepared for publication as books, computer programs, and advertising or marketing photographs.⁴¹ The Register of Copyrights has determined that these works are best for preregistration because of their “history of infringement prior to authorized commercial distribution.”⁴² To be available for preregistration, the work must be characterized as “being prepared for commercial distribution,” which requires “a reasonable expectation that the work will be commercially distributed to the public” and that “[p]reparation of the work has commenced and at least some portion of the work has been fixed in a tangible medium of expression.”⁴³

One commentator has suggested that the practical effect of preregistration is minimal: “In fact . . . most of the copyright cases that dealt with premature filings under the registration or application approaches would NOT be covered under the new preregistration system. Most did not involve works that were covered by the new preregistration regulations”⁴⁴ In fact, the commentator found only one case deciding the registration versus application approach issue in which the subject of the lawsuit would qualify for preregistration.⁴⁵

36. 17 U.S.C. § 410(d).

37. See *La Resolana Architects, PA v. Clay Realtors Angel Fire*, 416 F.3d 1195, 1204 n.9 (10th Cir. 2005), abrogated by *Reed Elsevier, Inc. v. Muchnick*, 130 S. Ct. 1237 (2010) (“As to the relation-back effect of registration, the effective date serves other purposes under the Act, such as the term of registration. It is not logical to assume that the relation-back provision subsumes the explicit requirements of § 411 and § 410(a).”).

38. 17 U.S.C. § 411(a).

39. Pub. L. No. 109-9, 119 Stat. 218 (2005) (codified in scattered sections of 17-18 U.S.C.).

40. 17 U.S.C. § 408(f)(4).

41. 37 C.F.R. § 202.16(b)(1) (2009).

42. 17 U.S.C. § 408(f)(2).

43. 37 C.F.R. § 202.16(b)(2)(i)-(ii).

44. Cain, *supra* note 17, at 389.

45. *Id.* at 390 (citing *Robinson v. Princeton Rev., Inc.*, 96 Civ. 4859 (LAK), 1996 U.S. Dist.

II. THE CIRCUIT SPLIT REGARDING ALTERNATE APPROACHES TO INTERPRETING SECTION 411(A)

Courts interpreting the Copyright Act's section 411(a) registration requirement are split between two opposing approaches: the "registration approach" and the "application approach."⁴⁶ As indicated previously, "[t]he dispute boils down to the issue of when a work is considered 'registered' for purposes of copyright law."⁴⁷ The registration approach employs a plain language view of Title 17, whereas the application approach implements an interpretation based more on policy and practicality.⁴⁸

A. The Registration Approach

Followed by the Tenth and Eleventh Circuits,⁴⁹ the registration approach interprets the registration requirement to mean that the work must actually be registered by the Copyright Office and that submission of an application alone is insufficient to satisfy 411(a) and allow filing of an infringement suit.⁵⁰ Because the Act distinguishes between application and registration and no language in the Act states that mere application suffices for registration, courts following the registration approach hold that registration is incomplete until the Register of Copyrights determines that copyright protection is warranted.⁵¹ In recently adopting the registration approach, the Tenth Circuit summarized its position:

[W]e reject the proposition that § 411 confers federal court jurisdiction for an infringement action upon mere submission of a copyright application to the Copyright Office. In our view, the statute requires more: actual registration by the Register of Copyrights. Until that happens, an infringement action will not lie in the federal courts.⁵²

These courts point to the plain text of section 411(a),⁵³ which gives no indication that application alone is sufficient:

[N]o civil action for infringement of the copyright in any United States

LEXIS 16932 (S.D.N.Y. Nov. 14, 1996)).

46. *La Resolana Architects, PA v. Clay Realtors Angel Fire*, 416 F.3d 1195, 1201-03 (10th Cir. 2005), *abrogated by* *Reed Elsevier, Inc. v. Muchnick*, 130 S. Ct. 1237 (2010).

47. *Iconbazaar, L.L.C. v. Am. Online, Inc.*, 308 F. Supp. 2d 630, 632 (M.D.N.C. 2004).

48. *La Resolana*, 416 F.3d at 1202-03.

49. *See id.* at 1202; *M.G.B. Homes, Inc. v. Ameron Homes, Inc.*, 903 F.2d 1486, 1488 (11th Cir. 1990).

50. *La Resolana*, 416 F.3d at 1202.

51. *See id.* at 1200-02.

52. *Id.* at 1205. Section 411(a) is no longer classified as a jurisdictional restriction, but as a precondition to filing copyright infringement suits. *See supra* notes 9-14 and accompanying text.

53. *See, e.g., id.* at 1200.

work shall be instituted until preregistration or registration of the copyright claim has been made in accordance with this title. In any case, however, where the deposit, application, and fee required for registration have been delivered to the Copyright Office in proper form and registration has been refused, the applicant is entitled to institute a civil action for infringement if notice thereof, with a copy of the complaint, is served on the Register of Copyrights.⁵⁴

In *Mays & Associates v. Euler*,⁵⁵ the United States District Court for the District of Maryland pointed out that the terms “application” and “registration” are used in the same section and clearly have different meanings.⁵⁶ The *Mays* court reasoned that section 411 lays out “the process of seeking registration . . . without labeling this process as *registration*,”⁵⁷ thus drawing a distinction between the process (application) and the result sought (registration). Another district court determined that there would be no reason for the Copyright Office to examine applications if these terms were synonymous.⁵⁸ Since the Act’s language “does not convey certain remedies and benefits upon application and other remedies and benefits upon registration,” the Tenth Circuit stated that the application approach would require a “topsy-turvy reading of Title 17 [of the Act].”⁵⁹

Section 410(a) of the Act specifies that “[w]hen, after examination, the Register of Copyrights determines that . . . the material deposited constitutes copyrightable subject matter . . . , the Register shall register the claim and issue to the applicant a certificate of registration under the seal of the Copyright Office.”⁶⁰ Some courts adopting the registration approach have argued that this language requires the Copyright Office to examine an application and then register a copyright claim; accordingly, they view the examination by the Copyright Office as a prerequisite to registration.⁶¹ Applying this reasoning, one court stated that the word registration “cannot possibly refer to the pre-examination receipt by the Copyright Office of the applicant’s fee, deposit, and application.”⁶²

54. 17 U.S.C. § 411(a) (2006 & Supp. 2009).

55. 370 F. Supp. 2d 362 (D. Md. 2005).

56. *Id.* at 368.

57. *Id.*

58. *Robinson v. Princeton Rev., Inc.*, No. 96 Civ. 4859 (LAK), 1996 U.S. Dist. LEXIS 16932, at *21 (S.D.N.Y. Nov. 14, 1996).

59. *La Resolana Architects, PA v. Clay Realtors Angel Fire*, 416 F.3d 1195, 1204 (10th Cir. 2005), *abrogated by Reed Elsevier, Inc. v. Muchnick*, 130 S. Ct. 1237 (2010).

60. 17 U.S.C. § 410(a) (2006).

61. See, e.g., *Ryan v. Carl Corp.*, No. C 97-3873 FMS, 1998 U.S. Dist. LEXIS 9012, at *5-7 (N.D. Cal. June 15, 1998) (“Because . . . [section 410] indicates that the Copyright Office, not the applicant, registers a claim, and that examination is a prerequisite to registration, the section cuts against plaintiffs’ position of automatic registration [through application only].”).

62. *Loree Rodkin Mgmt. Corp. v. Ross-Simons, Inc.*, 315 F. Supp. 2d 1053, 1055 (C.D. Cal. 2004).

The Tenth Circuit also argued in *La Resolana Architects, PA v. Clay Realtors Angel Fire* that two amendments to the Copyright Act since its enactment in 1976 support the registration approach.⁶³ First, Congress passed the Berne Convention Implementation Act in 1988 to “ally the United States with a set of international rules and regulations.”⁶⁴ While considering the Berne Act, which eliminated many previously existing formalities to copyright protection, Congress considered amending 411(a) to read “registration is not a prerequisite to the institution of a civil action for infringement of copyright.”⁶⁵ The House of Representatives rejected the proposed modification, instead amending 411(a) to allow only foreign authors to initiate suit without first registering.⁶⁶ Congress clarified that “[r]egistration is continued as a prerequisite to suit by *domestic* authors. Only foreign origin works are excepted from the registration requirement.”⁶⁷ The Tenth Circuit explained that this outcome “confirms our view of the Act: copyrights that originate in the United States must be registered before the owner can sue for infringement.”⁶⁸ Second, the Tenth Circuit pointed to the 2005 FECA amendment, which allows for preregistration of certain works.⁶⁹ The court argued that this amendment strengthened its registration approach position:

[T]he adoption of FECA further confirms our statutory analysis. Indeed, the availability of a preregistration scheme would in whole or in part address the problem presented by this case: the need to sue for infringement to prevent dilution of a copyright but the inability to do so without completed registration.⁷⁰

1. *The Certificate.*—Some courts following the registration approach additionally require an applicant to receive a paper certificate of registration from the Copyright Office before filing an infringement action.⁷¹ The Tenth Circuit, however, specifically stated in *La Resolana* that it does not require actual receipt of the certificate because the statutory language does not seem to require it and because of the delay between registration and the owner’s receipt of the certificate.⁷² Nonetheless, a Minnesota district court in *Tri-Marketing v.*

63. 416 F.3d at 1205.

64. *Id.*

65. *Id.* at 1205-06 (quoting S. REP. NO. 100-352, at 46 (1998), reprinted in 1988 U.S.C.C.A.N. 3706, 3743).

66. *Id.* at 1206.

67. *Id.* (quoting 134 CONG. REC. H10095).

68. *Id.* Note that application approach courts would agree that registration is required but would argue that it is satisfied by submission of an application. *See infra* Part II.B.

69. *Id.* at 1206-07; *see supra* notes 39-45 and accompanying text for discussion of FECA.

70. *Id.* at 1207.

71. *See, e.g., Strategy Source, Inc. v. Lee*, 233 F. Supp. 2d 1, 3-4 (D.D.C. 2002) (“[P]ermitting an infringement lawsuit to go forward in the absence of a registration certificate or denial of the same is in tension with the language of section 411(a) of the Copyright Act.”).

72. *La Resolana*, 416 F.3d at 1202-03 (noting that “a court’s jurisdiction does not turn on

Mainstream Marketing Services, Inc.,⁷³ citing to *La Resolana*, recently stated broadly that the registration approach requires receipt of a certificate as a prerequisite to an infringement lawsuit.⁷⁴ Ironically, this universal statement that receipt of a certificate is required is inaccurate precisely because of *La Resolana*. The *Tri-Marketing* court's citation to *La Resolana*, in which the Tenth Circuit explicitly stated that it does not require a certificate, is one of numerous examples of the confusion among federal courts regarding the copyright registration issue.

2. *Flawed Names for the Registration Approach*.—The “registration approach,” as referred to in this Note, is referred to by some commentators as the “issuance approach”⁷⁵ and by others as the “approval approach.”⁷⁶ Yet each of these terms has its problems. First, “issuance” has a connotation that the Copyright Office must actually send something out or “issue” a copyright. According to the Tenth Circuit in *La Resolana*, however, the Copyright Office need only have approved or rejected an application, and the owner need not receive anything issued by the Copyright Office.⁷⁷ Since the Tenth Circuit—perhaps the foremost proponent of the registration approach—does not require the copyright owner to receive any issuance, referring to this position as the “issuance approach” is undoubtedly misleading. “Approval” is more accurate in this sense; however, this name is also nonetheless imprecise because an owner may sue to enforce a rejected copyright, not just one that has been approved.⁷⁸ A more fitting title, one that would most accurately describe what these courts require to satisfy the section 411(a) prerequisite, would be the “approval or denial” approach. Despite its technical accuracy, this name is not very catchy and appears at first glance to be an oxymoron since it includes opposite results. It is thus doubtful that any esteemed federal judge or commentator will adopt this moniker anytime soon.

The name “registration,” while most commonly used to describe the Tenth Circuit’s approach, suffers from the same defect as the name “approval” in that a copyright that has been refused registration may still be sued upon.⁷⁹ In addition, use of “registration” is confusing since the meaning of the word “registration” in the Act and the time when registration is accomplished are at the center of the debate among the circuit courts. Neither approach argues that “registration,” as required by section 411(a), is unnecessary;⁸⁰ the view of the

the existence of a paper certificate, but rather on the fact of registration, however it is demonstrated” and that “such registration occurs when the Copyright Office approves the application.”).

73. Civ. No. 09-13 (DWF/RLE), 2009 U.S. Dist. LEXIS 42694 (D. Minn. May 19, 2009).

74. *Id.* at *6-7.

75. See Bracey, *supra* note 7, at 127.

76. See Hogan, *supra* note 16, at 854.

77. *La Resolana*, 416 F.3d at 1202-03.

78. 17 U.S.C. § 411(a) (2006 & Supp. 2009); *see also supra* text accompanying note 32.

79. 17 U.S.C. § 411(a); *see also supra* text accompanying note 32.

80. See, e.g., *La Resolana*, 416 F.3d at 1205; *Forasté v. Brown Univ.*, 248 F. Supp. 2d 71, 76 (D.R.I. 2003) (“Copyright registration is a condition precedent and a jurisdictional prerequisite to

“application” approach is that this statutorily required “registration” occurs when a copyright owner applies for registration.⁸¹ Thus, using the “registration” label titles one interpretation of the issue after the source of the issue itself—the meaning of “registration.” Perhaps some of the confusion among federal courts arises from failing to differentiate the general statutory requirement of “registration” in section 411(a)—whatever it may mean—from the one interpretation of the requirement which holds that the Copyright Office must actually examine and choose to register the copyright (or deny registration).⁸²

B. The Application Approach

Followed primarily by the Fifth Circuit,⁸³ the application approach contends that a copyright owner satisfies the registration requirement by submitting a copyright application.⁸⁴ The application approach essentially says that for policy and practicality reasons, the “registration” required by 411(a) as a precondition to suit does not mean *actual* registration by the Copyright Office.⁸⁵ Rather, the application approach holds that application for such actual registration fulfills the statutory “registration” requirement.⁸⁶ Courts following the application approach point to section 410(d) to justify their position: “*The effective date of a copyright registration is the day on which an application, deposit, and fee, which are later determined by the Register of Copyrights or by a court of competent jurisdiction to be acceptable for registration, have all been received in the Copyright Office.*”⁸⁷ They read this provision to mean that the effective date is effective immediately, not retroactively after the Copyright Office makes a decision.⁸⁸ The court in *Forasté v. Brown University* mentioned that registration approach courts interpret section 410(d) “to mean that registration is consummated only after an application is examined, considered, and accepted by the Copyright Office, and is then ‘backdated’ to the time the application is received” but that “[the registration approach interpretation] ignores the statute’s mandate that the merits of the application materials are ‘later determined,’ that is, determined at some

the filing of an infringement action.”).

81. *Forasté*, 248 F. Supp. 2d at 77 (“[R]egistration occurs *on the day* the Copyright Office receives all of the necessary application materials . . .”).

82. See, e.g., *La Resolana*, 416 F.3d at 1205; *Apple Barrel Prods., Inc. v. Beard*, 730 F.2d 384, 386-87 (5th Cir. 1984).

83. *Positive Black Talk, Inc. v. Cash Money Records, Inc.*, 394 F.3d 357, 365 (5th Cir. 2004), abrogated by *Reed Elsevier, Inc. v. Muchnick*, 130 S. Ct. 1237 (2010); *Apple Barrel*, 730 F.2d at 387-88.

84. *Forasté*, 248 F. Supp. 2d at 77.

85. See *La Resolana*, 416 F.3d at 1203.

86. See *Apple Barrel*, 730 F.2d at 386-87 (noting that federal jurisdiction is satisfied by “payment of the required fee, deposit of the work in question, and receipt by the Copyright Office of a registration application.”).

87. See 17 U.S.C. § 410(d) (2006); *Forasté*, 248 F. Supp. 2d at 77.

88. See *Forasté*, 248 F. Supp. 2d at 77 n.10.

time after the right to sue comes into being.”⁸⁹ In other words, application approach courts interpret the “effective date” provision in section 410(d) to mean that a copyright is effectively registered immediately when the application materials are submitted, with the effective date being that same day. According to the application approach courts’ interpretation, the copyright owner has established the right to sue on this effective date, and whether the Copyright Office decides to accept or reject the application is an issue for a separate inquiry which does not affect the owner’s right to sue.⁹⁰

A leading treatise on copyright law, *Nimmer on Copyright* (“Nimmer”), supports the application approach.⁹¹ Nimmer states that “[a] party who seeks to register may proceed to litigate a claim, regardless of whether the Copyright Office ultimately issues the certificate, or by contrast denies it. Accordingly, it makes sense under the 1976 Act to refer to *application for registration* as a condition to filing an infringement action . . .”⁹² Nimmer further argues that the registration approach results from “hyper-technical application” and that policy rationales support the application approach.⁹³

The policy rationales mentioned in *Nimmer* are indeed central to the reasoning of application approach proponents. Courts adopting the application approach argue that delaying an infringement claim until the claimant has received a response from the Copyright Office is senseless since the claimant will be able to sue regardless of whether his application is ultimately granted or rejected.⁹⁴ The application approach also avoids the inefficient situation where an infringement claim is dismissed, then re-filed shortly thereafter following a decision by the Copyright Office.⁹⁵ Accordingly, one district court stated that “[t]o best effectuate the interests of justice and promote judicial economy, the court endorses the position that a plaintiff may sue once the Copyright Office receives the plaintiff’s application, work, and filing fee.”⁹⁶ In addition, since applications are typically pending in the Copyright Office for a number of months,⁹⁷ allowing copyright owners to bring a claim after applying allows them to prevent their work from being infringed while waiting for approval. Furthermore, the availability of immediate copyright enforcement under the

89. *Id.*

90. This argument is buffered by the fact that the separate inquiry of whether the work is suitable for registration does not affect the owner’s right to sue because the Act establishes an owner’s right to sue if his or her application is rejected. *See* 17 U.S.C. § 411(a).

91. *See generally* 2 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT § 7.16 (2009).

92. *Id.* § 7.16(B)(1)(a)(i).

93. *Id.*

94. *See, e.g.*, Apple Barrel Prods., Inc. v. Beard, 730 F.2d 384, 386-87 (5th Cir. 1984).

95. Bracey, *supra* note 7, at 141.

96. Int’l Kitchen Exhaust Cleaning Ass’n. v. Power Washers of N. Am., 81 F. Supp. 2d 70, 72 (D.D.C. 2000).

97. *See* U.S. COPYRIGHT OFFICE, *supra* note 35.

application approach strengthens the “incentive to create.”⁹⁸

III. CONFUSION IN THE SEVENTH AND EIGHTH CIRCUITS

The lack of clarity among federal courts that have faced the application versus registration issue may be best illustrated by case law from the Seventh and Eighth Circuits. These cases exemplify the importance of precise language in opinions dealing with the application versus registration issue. In both circuits, opinions that appeared to choose an approach have been met with varied responses and confusion among their respective district courts.⁹⁹ By a close examination of the relevant cases in these circuits, this Note attempts to identify the reasons for confusion and the possible causes of seemingly inconsistent judicial language.

A. *Seventh Circuit*

1. *Chicago Board of Education v. Substance, Inc.*—In *Chicago Board of Education v. Substance, Inc.*,¹⁰⁰ a Chicago public school teacher published six Chicago Academic Standards Exams in order to demonstrate that they were poor tests.¹⁰¹ The school board had expended in excess of one million dollars to create the entire series of forty-four exams.¹⁰² The school board’s suit alleged that the teacher’s actions impaired the exams’ value because the exams were intended to be kept secret to allow reuse of questions.¹⁰³ The teacher argued that the school board lacked a valid copyright registration on which to sue.¹⁰⁴

In response to the teacher’s claim, Judge Posner first appeared to adopt an application approach: “Although a copyright no longer need be registered with the Copyright Office to be valid, *an application for registration must be filed* before the copyright can be sued upon.”¹⁰⁵ The first part of this sentence—stating that “a copyright no longer need be registered with the Copyright Office to be valid”—is a simple statement of the fact that the law recognizes copyright ownership without registration;¹⁰⁶ it does not refer to whether the owner may sue. However, the second clause—stating that “an application for registration must be filed before the copyright can be sued upon”—certainly seems to indicate that actual registration is not required, but rather, that filing of an application will suffice. In addition, Judge Posner cited to *Nimmer*, the treatise widely recognized as advocating the application

98. Bracey, *supra* note 7, at 141.

99. See *infra* notes 100-53 and accompanying text.

100. 354 F.3d 624 (7th Cir. 2003).

101. *Id.* at 625-27.

102. *Id.* at 626.

103. *Id.* at 628.

104. *Id.* at 631.

105. *Id.* (emphasis added) (citing 17 U.S.C. § 411(a) (2006 & Supp. 2009)).

106. See 17 U.S.C. § 102(a) (“Copyright protection subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression . . .”).

approach, as support for his assertion that “an application for registration must be filed before the copyright can be sued upon.”¹⁰⁷ He cited to the same section of *Nimmer* cited to by the Fifth Circuit in *Apple Barrel Productions, Inc. v. Beard* in support of its application approach.¹⁰⁸

It is possible that Judge Posner was unaware either of the circuit split or that his statement and citation to *Nimmer* would give the impression that he was choosing one side (the application approach). Perhaps he looked to *Nimmer* to find a discussion of the 411(a) requirement but did not mean to reference *Nimmer*’s preferred position on the split. Or perhaps Judge Posner only meant to say that filing an application for registration is necessary—a statement with which registration approach proponents would not disagree—but not that it is sufficient. Whether application alone is sufficient is where the two views differ. Regardless of Judge Posner’s thoughts while writing the *Chicago Board* opinion, the statement that “an application for registration must be filed before the copyright can be sued upon” is, at the least, confusing and misleading if indeed he did not intend to advocate the application approach.

The *Chicago Board* opinion goes on to explain that the school board had applied for registration of the exams and that the Copyright Office did indeed register the copyright.¹⁰⁹ Judge Posner then engaged in a discussion that would only be relevant under a registration approach:

Had the claim [that the school board owned copyrights in the tests] been false, the registration should not have issued and maybe therefore the copyright could not have been sued upon. . . . Or maybe yes, because the copyright would have been registered, and because the statute requires only a refused registration, which might be the equivalent of an improper registration, not an actual registration, as the premise for the suit. We need not decide [because the board’s claim that it owned registered copyrights in the tests was true].¹¹⁰

If Judge Posner had previously meant to say that only application is required, there would have been no reason for discussing the effects of the copyright being registered (albeit improperly) or denied. In contrast to the earlier sentence in the opinion that seemed to be espousing the application approach, this later statement appears as if Judge Posner was talking about the registration approach, where mere application would not be enough to allow a copyright to be sued upon.

2. *Gaiman v. McFarlane*.—In *Gaiman v. McFarlane*,¹¹¹ Judge Posner wrote again for the Seventh Circuit and explained that “[t]he significance of registration is that it is a prerequisite to a suit to enforce a copyright. More precisely, an

107. *Chi. Bd. of Educ.*, 354 F.3d at 631 (citing NIMMER & NIMMER, *supra* note 91, § 7.16[B][1][a]); see *supra* text accompanying notes 91-93 (discussing *Nimmer*’s position).

108. See *Chi. Bd. of Educ.*, 354 F.3d at 631; *Apple Barrel Prods., Inc. v. Beard*, 730 F.2d 384, 386-87 (5th Cir. 1984).

109. *Chi. Bd. of Educ.*, 354 F.3d at 631.

110. *Id.*

111. 360 F.3d 644 (7th Cir. 2004).

application to register must be filed, *and either granted or refused*, before suit can be brought.”¹¹² As in *Chicago Board*, the court was not applying this statement to a situation where it had to choose between the two approaches in order to decide the case because McFarlane had successfully registered his copyrights.¹¹³ However, *Gaiman* is still informative regarding the Seventh Circuit’s position on the issue, especially since it was written by Judge Posner less than two months after he penned the *Chicago Board* opinion.¹¹⁴

Immediately following this brief discussion of registration’s significance, Judge Posner wrote another sentence that provides further insight into *Chicago Board*: “There is an interesting question, left open in our recent decision in [*Chicago Board*] . . . and unnecessary to decide in this case either, whether if registration is granted by mistake the registrant may nonetheless sue.”¹¹⁵ Again, if *Chicago Board* had intended to adopt an application approach, it would be irrelevant whether registration was later granted (even if by mistake) or refused. The plaintiff would be able to sue as soon as the Copyright Office received his application, and the hypothetical Judge Posner added as an aside here would be irrelevant. This comment is thus strong evidence of Judge Posner’s own interpretation of his *Chicago Board* opinion, fresh in his mind from less than two months prior. Even more significantly, his clear statement that application must be “either granted or refused” appears to be a strong indication that the Seventh Circuit supports the traditional plain-reading registration approach to interpreting section 411(a).¹¹⁶

3. *Goss International Americas, Inc. v. A-American Machine & Assembly Co.*—Despite the apparent clarity of the excerpt from *Gaiman*, the United States District Court for the Northern District of Illinois cited only to *Chicago Board* in 2007 when deciding the application versus registration issue in *Goss International Americas, Inc. v. A-American Machine & Assembly Co.*¹¹⁷ After plaintiff Goss had applied for copyright registration, defendant A-American Machine & Assembly posted Goss’s diagrams on its website without his permission.¹¹⁸ The court determined that Goss’s drawings were entitled to copyright protection under the Act and recognized that it was squarely facing the issue on which the circuits are split—whether the copyright could be sued upon following application but prior to a decision from the Copyright Office.¹¹⁹ The court quoted the statement from *Chicago Board* that “an application for registration must be filed before the copyright can be sued upon.”¹²⁰ Oddly, it

112. *Id.* at 654-55 (emphasis added) (citing 17 U.S.C. § 411(a)(2006 & Supp. 2009)).

113. *Id.* at 654.

114. *See id.* (decided Feb. 24, 2004); *Chi. Bd. of Educ.*, 354 F.3d 624 (decided Dec. 31, 2003).

115. *Gaiman*, 360 F.3d at 655.

116. *See id.* at 654-55.

117. No. 07 C 3248, 2007 U.S. Dist. LEXIS 88382, at *6 (N.D. Ill. Nov. 30, 2007).

118. *Id.* at *2.

119. *Id.* at *5.

120. *Id.* at *6 (quoting *Chi. Bd. of Educ. v. Substance, Inc.*, 354 F.3d 624, 631 (7th Cir. 2003)).

attributed the application approach's "effective date" argument to Judge Posner,¹²¹ even though Judge Posner never actually made this point: "Judge Posner bases this statement on the fact that once the Copyright Office approves an application, it retroactively lists as the effective date of the registration the date on which it received all application materials from the applicant."¹²² This attempt to divine the reasoning behind Judge Posner's unclear *Chicago Board* opinion is essentially guesswork, but it indicates that the *Goss* court understood *Chicago Board* to be advocating an application approach. Not only did Judge Posner not make the "effective date" argument for an application approach, but it is far from clear whether he even meant to endorse this approach at all, especially in light of his comments in *Gaiman*.

The *Goss* court noted that the *Chicago Board* case did not force the Seventh Circuit to choose an approach:

[*Chicago Board*] differs somewhat from the instant case, though, because in the case before Judge Posner the Copyright Office had issued a registration certificate. In the instant case, plaintiff has been waiting since May 2007 for some action by the Copyright Office. In the meantime, plaintiff has allegedly suffered financial damage as a result of defendant's copying of its images. . . .¹²³

Without mentioning *Gaiman*, the court followed the application approach of the Fifth Circuit.¹²⁴ The *Goss* court's explanation exemplifies classic application approach reasoning, based on the right of a copyright owner to sue after the Copyright Office has refused to register his application:

[Because the Copyright Act] allows a party to sue for infringement not only after approval of a copyright registration application, but also after a refusal of that registration application . . . [i]t hardly seems fair, then, to refuse plaintiff the opportunity to sue for infringement until the Copyright Office takes some action when plaintiff will be able to sue after that determination, regardless of the outcome.¹²⁵

Although the district court in *Goss* recognized that *Chicago Board* had not faced the precise situation that would require choosing one approach over the other, it took note of and appears to have been influenced by what it perceived as Judge Posner's support of the application approach. The *Goss* court did not recognize that the language following the quoted section from *Chicago Board* makes uncertain which approach Judge Posner meant to advocate, if he intended to

121. See *supra* text accompanying notes 87-90 (discussing the effective date argument based on section 410(d)).

122. *Goss*, 2007 U.S. Dist. LEXIS 88382, at *6.

123. *Id.* at *6-7.

124. *Id.* at *7. The court actually stated that it adopted the reasoning of the Fifth and Eighth Circuits in choosing the application approach. However, the position of the Eighth Circuit is not clearly established, as discussed *infra*, Part III.B.

125. See *id.* at *7.

advocate one at all. Nor did it consider what appears to be much more persuasive dicta from *Gaiman*, in which the Seventh Circuit—with Judge Posner again writing—added that the application must be either granted or refused before the owner can sue on the copyright.¹²⁶

4. *Woollen v. Indianapolis-Marion County Public Library*.—The United States District Court for the Southern District of Indiana stated in *Woollen v. Indianapolis-Marion County Public Library*,¹²⁷ regarding the two approaches, that “[t]he Seventh Circuit has not yet addressed this issue directly, although one decision suggests that this circuit may favor the ‘application approach.’”¹²⁸ The court discussed the Seventh Circuit’s language from *Chicago Board* that “an application for registration must be filed before the copyright can be sued upon” but noted that “[t]he comment in *Chicago Board of Education*, however, was made in passing.”¹²⁹ The *Woollen* court was able to avoid deciding the case because the Copyright Office registered the plaintiff’s copyright while the defendant’s motion to dismiss was pending.¹³⁰ The court allowed the plaintiff to amend its complaint to assert that its copyright had been registered and a certificate had been issued,¹³¹ thus satisfying section 411(a) even under the stricter registration approach. The court noted that the door was left open by the Seventh Circuit as to when federal courts obtain jurisdiction over copyright infringement suits, stating that “this court acquired jurisdiction to hear its claims, at least from the date of the Copyright Office’s registration and perhaps sooner.”¹³²

The *Goss* and *Woollen* opinions indicate that district courts have not understood *Gaiman* as the Seventh Circuit’s adoption or promotion of the registration approach despite the seemingly clear language in that opinion. Commentaries also do not attribute either the registration or the application approach to the Seventh Circuit.¹³³ Absent a direct determination by the Seventh Circuit, the application approach adopted in *Goss* appears to be the leading persuasive authority for the Northern District of Illinois.

5. *Brooks-Ngwenya v. Indianapolis Public Schools*.—When the Seventh Circuit was recently presented with another opportunity to explain its prior holdings, it chose not to do so in *Brooks-Ngwenya v. Indianapolis Public Schools*.¹³⁴ The court noted that “[t]he circuits have split over whether registration is complete when an application is made or only after the Copyright

126. See *Gaiman v. McFarlane*, 360 F.3d 644, 654-55 (7th Cir. 2004).

127. No. 1:06-cv-0662-JDT-TAB, 2006 U.S. Dist. LEXIS 52788 (S.D. Ind. July 28, 2006).

128. *Id.* at *6.

129. *Id.*

130. *Id.* at *7.

131. *Id.* at *8.

132. *Id.* at *9. Section 411(a) is no longer classified as a jurisdictional restriction, but as a precondition to filing copyright infringement suits. See *supra* notes 9-14 and accompanying text.

133. See, e.g., Joshua P. Graham, *Graham on Goss International Americas, Inc. v. A-American Machine & Assembly*, 2008 EMERGING ISSUES 1757 (2008).

134. 564 F.3d 804 (7th Cir. 2009).

Office has acted on the application.”¹³⁵ At the end of a string cite to courts on either side of the split, the court included “[compare to] *Chicago Board of Education v. Substance, Inc.* . . . (‘an application for registration must be filed before the copyright can be sued upon’).”¹³⁶ The court did not discuss its prior *Chicago Board* decision besides listing it in this string cite, apparently indicating that *Chicago Board* was not meant to take a position. As in *Chicago Board* and *Gaiman*, the facts in *Brooks-Ngwenya* did not require the court to take a position on whether application alone satisfies the section 411(a) requirement because the Copyright Office had denied the plaintiff’s application for registration.¹³⁷ *Brooks-Ngwenya* seems to clarify that despite indications to the contrary, the Seventh Circuit has not endorsed either the application or registration approach.

B. Eighth Circuit

1. Action Tapes, Inc. v. Mattson.—In *Action Tapes, Inc. v. Mattson*,¹³⁸ the Eighth Circuit stated that “the copyright owner may not sue for infringement under the federal Copyright Act until the owner has delivered ‘the deposit, application, and fee required for registration’ to the United States Copyright Office, a branch of the Library of Congress.”¹³⁹ Perhaps all the court intended by this sentence was to paraphrase section 411(a). Instead, however, the court seemed to advocate an application approach by stating that a copyright owner must deliver the requisite application materials before suing.¹⁴⁰ This statement says nothing about needing to receive a response from the Copyright Office. Like the Seventh Circuit’s language in *Chicago Board* that “an application for registration must be filed before the copyright can be sued upon,”¹⁴¹ however, this statement raises the “necessary vs. sufficient” distinction. The Eighth Circuit in *Action Tapes* may have intended to say only that delivering the application materials is a necessary step towards fulfilling the 411(a) prerequisite without meaning to suggest that this is all a copyright owner must do (sufficient). There would be no opposition to the statement that application, at the very least, is a necessary prerequisite to an infringement suit. The Eighth Circuit may have been simply stating this non-controversial rule without intending to hold that application alone is sufficient. As far as the issue of the 411(a) prerequisite related to the case at bar, all that mattered was that application for copyright protection was necessary.¹⁴² Since Action Tapes had not properly applied for a computer program copyright, section 411(a) barred its infringement suit.¹⁴³

135. *Id.* at 806.

136. *Id.*

137. *Id.*

138. 462 F.3d 1010 (8th Cir. 2006).

139. *Id.* at 1013 (quoting 17 U.S.C. § 411(a) (2006 & Supp. 2009)).

140. *See id.* (quoting 17 U.S.C. § 411(a)).

141. Chi. Bd. of Educ. v. Substance, Inc., 354 F.3d 624, 631 (7th Cir. 2003).

142. *See Action Tapes*, 462 F.3d at 1013.

143. *Id.*

Whatever the Eighth Circuit intended in *Action Tapes*, following that decision, courts began to list the Eighth Circuit along with the Fifth Circuit as the two circuits that had adopted the application approach.¹⁴⁴ Clearly, the Eighth Circuit's statement in *Action Tapes* that "the copyright owner may not sue for infringement . . . until the owner has delivered 'the deposit, application, and fee required for registration'" was widely understood as an adoption or at least a strong endorsement of the application approach.¹⁴⁵ This understanding that the Eighth Circuit adopted the application approach in *Action Tapes* is precisely what the copyright owner relied on in the recent 2009 district court case of *Tri-Marketing v. Mainstream Marketing Services*.¹⁴⁶

2. *Tri-Marketing v. Mainstream Marketing Services*.—In *Tri-Marketing v. Mainstream Marketing Services*,¹⁴⁷ TRI applied for registration of two versions of its website,¹⁴⁸ but it had not received any response from the Copyright Office before instituting its lawsuit.¹⁴⁹ TRI pointed to *Action Tapes* as the Eighth Circuit's adoption of the application approach, arguing that *Action Tapes* "squarely addresse[d]" the issue.¹⁵⁰ The court was not persuaded:

While the language in *Action Tapes* appears to support TRI's position in this case, the Eighth Circuit in *Action Tapes* was not presented with the precise jurisdictional issue before this Court, specifically whether complying with copyright application requirements satisfies the jurisdictional requirements under § 411(a). Accordingly, the language

144. See *Brooks-Ngwenya v. Indianapolis Pub. Sch.*, 564 F.3d 804, 806 (7th Cir. 2009) ("The circuits have split over whether registration is complete when an application is made or only after the Copyright Office has acted on the application. Compare *Action Tapes* . . . (application is sufficient)"); *DO Denim, LLC v. Fried Denim, Inc.*, No. 08 Civ. 10947, 2009 U.S. Dist. LEXIS 51512, at *5-6 (S.D.N.Y. June 17, 2009) ("Compare *La Resolana* . . . ('the Register of Copyrights must affirmatively determine copyright protection is warranted . . . before registration occurs under the Act. . . .') with *Action Tapes, Inc. v. Mattson* . . . ('the copyright owner may not sue for infringement under the federal Copyright Act until the owner has delivered the deposit, application, and fee required for registration to the United States Copyright Office')"); *Goss Int'l Ams., Inc. v. A-Am. Mach. & Assemb. Co.*, No. 07 C 3248, 2007 U.S. Dist. LEXIS 88382, at *5-6 (N.D. Ill. Nov. 30, 2007) ("The Fifth and Eighth Circuits, however, hold that a plaintiff need only send the application, deposit, and fee to the Copyright Office to file an infringement claim."); *Kaye Homes, Inc. v. Original Custom Homes Corp.*, No. 2:07-cv-392-FtM-29SPC, 2007 U.S. Dist. LEXIS 60847, at *5-6 (M.D. Fla. Aug. 20, 2007) ("The Circuit Courts of Appeal have split in their interpretation of this statute, with one group finding that filing an application is sufficient to vest jurisdiction (citing *Action Tapes*, 462 F.3d at 1010))).

145. See *Action Tapes*, 462 F.3d at 1013 (quoting 17 U.S.C. § 411(a) (2006)).

146. Civ. No. 09-13 (DWF/RLE), 2009 U.S. Dist. LEXIS 42694 (D. Minn. May 19, 2009).

147. *Id.*

148. *Id.* at *2.

149. *Id.* at *6.

150. *Id.* at *8.

relied on by TRI is persuasive but not binding on this [c]ourt.¹⁵¹

Although the court determined that it was not technically bound by *Action Tapes*, this persuasive precedent, along with “the interests of justice and judicial economy,” led the court to endorse the application approach.¹⁵²

As copyright law in the Eighth Circuit now stands, *TRI Marketing* and the persuasive dicta from *Action Tapes* support an application approach, whereas older district court cases support the registration approach.¹⁵³ Another district court within the Eighth Circuit could simply choose not to follow *TRI Marketing*, however, and dismiss a case in which the Copyright Office has not registered the owner’s copyright by the time the suit is filed. The Eighth Circuit should look for an opportunity to clarify its intent for an application approach—if that was indeed its intent—so that all district courts within the circuit will be applying the same rule.

IV. PROPOSAL

A. Absent Congressional Amendment, Courts Should Follow the Registration Approach

As currently written, the Copyright Act by its plain language prohibits suit over a copyright after application but before a response from the Copyright Office. Simply stated, the application approach ignores the congressionally-intended plain meaning of section 411(a).¹⁵⁴ Courts should look to the Tenth Circuit’s well-reasoned opinion in *La Resolana* and follow the registration approach based on the plain meaning evident in the statute, as it has been drafted and enacted by Congress.

The law is well established that “[i]t is the province of the legislature to make the laws, and of the court to enforce them.”¹⁵⁵ The courts’ role is not to “fix” statutes that they may not see as the best policy by inventing fictions—such as that “registration” does not really mean registration, as application approach courts are essentially doing.¹⁵⁶ The Tenth Circuit began its analysis in *La Resolana* by noting that “we start with the language of the statute. If the statutory language is not ambiguous, and the ‘statutory scheme is coherent and consistent,’ our inquiry ends.”¹⁵⁷ A simple reading of section 411(a) reveals that

151. *Id.* at *8-9. Section 411(a) is no longer classified as a jurisdictional restriction, but as a precondition to filing copyright infringement suits. *See supra* notes 9-14 and accompanying text.

152. *Id.* at *10.

153. *See, e.g.*, Denenberg v. Berman, No. 4:02CV7, 2002 U.S. Dist. LEXIS 20490, at *4-5 (D. Neb. July 23, 2002); Proulx v. Hennepin Tech. Ctrs. Dist. No. 287, Civ. No. 4-79-637, 1981 U.S. Dist. LEXIS 17634, at *4 (D. Minn. Dec. 7, 1981).

154. *See infra* note 161 and accompanying text.

155. *Barrett v. State*, 229 U.S. 26, 30 (1913).

156. *See supra* Part II.B.

157. *La Resolana Architects, PA v. Clay Realtors Angel Fire*, 416 F.3d 1195, 1200 (10th Cir. 2005), *abrogated by* *Reed Elsevier, Inc. v. Muchnick*, 130 S. Ct. 1237 (2010) (quoting Barnhart

Congress used both “application” and “registration” and gave the two terms different meanings.¹⁵⁸ Thus, section 411(a)’s disallowance of a copyright infringement action before “preregistration or registration” conveys to the reader that “application,” discussed in the same subsection, is insufficient to allow an action for copyright infringement.¹⁵⁹ Otherwise, the next sentence, allowing suit where “the deposit, application, and fee required for registration have been delivered to the Copyright Office in proper form and registration has been refused,”¹⁶⁰ would be meaningless. Although section 411(a) creates an odd result by allowing a suit whether the Copyright Office grants or refuses registration, it makes clear that the right to sue comes into being upon such grant or refusal and not upon the preliminary step of application. The United States District Court for the District of Columbia, following the registration approach, wrote that “[t]o conclude that registration or the refusal of registration by the Copyright Office are not . . . prerequisites is to disregard the plain language of these statutes and to in effect re-write them, which . . . this [c]ourt cannot do.”¹⁶¹

It is unlikely that any application approach courts read section 411(a) and truly come away with the understanding that the statute says application is sufficient. Section 411(a) is not an ambiguous statute that is open to multiple interpretations. No application approach court has been so bold as to claim that Congress specified in the Act that an owner’s application for copyright protection alone allows the owner to bring suit.¹⁶² Rather, the application approach simply ignores the clear meaning of this provision for policy reasons. However, as the Tenth Circuit recognized, “[w]hatever the practical force of this argument . . . [courts] cannot ignore the plain meaning of the statute, nor change the legislative scheme.”¹⁶³

B. Congress Should Amend the Act to Allow Infringement Suits After Application

Some measure of clarity is needed with respect to section 411(a) since the registration and application approaches are currently applied seemingly at random, as each court sees fit.¹⁶⁴ Copyright holders outside of the Fifth, Tenth, and Eleventh Circuits—those in which the circuit courts have definitively ruled

v. Sigmon Coal Co., 534 U.S. 438, 450 (2002)).

158. See 17 U.S.C. § 411(a) (2006 & Supp. 2009); Mays & Assocs., Inc. v. Euler, 370 F. Supp. 2d 362, 368 (D. Md. 2005).

159. See 17 U.S.C. § 411(a).

160. See *id.*

161. Strategy Source, Inc. v. Lee, 233 F. Supp. 2d 4, 10 (D.D.C. 2002).

162. Rather, application approach courts argue that policy reasons support such an interpretation by the judiciary. See *supra* Part II.B.

163. La Resolana Architects, PA v. Clay Realtors Angel Fire, 416 F.3d 1195, 1204 (10th Cir. 2005), abrogated by Reed Elsevier, Inc. v. Muchnick, 130 S. Ct. 1237 (2010).

164. See Brooks-Ngwenya v. Indianapolis Pub. Sch., 564 F.3d 804, 806 (7th Cir. 2009) (listing opinions on each side of the split).

on the application versus registration issue—have little idea what prerequisites they must satisfy before initiating suit. The glaring lack of certainty and uniformity regarding section 411(a)'s registration requirement is confusing copyright holders and courts alike¹⁶⁵ and can be best remedied by clarification from Congress. Notwithstanding the courts' duty to follow the registration approach in observance of the Act's plain meaning, the various courts and commentators that have advocated the application approach are indeed supported by significant policy rationales.¹⁶⁶ Congress should take note of the reasoning in support of the application approach and amend the Copyright Act to allow copyright holders to institute infringement suits after they have delivered their application materials to the Copyright Office for review.¹⁶⁷

The most persuasive point favoring the application approach is that because an applicant will be able to sue eventually regardless of whether his application is granted or rejected,¹⁶⁸ it makes little sense to require him to wait for a response.¹⁶⁹ Requiring owners to wait until registration has been completed forces the owner-applicant whose copyright is being infringed upon to sit by idly after applying, often for more than a year,¹⁷⁰ before the Copyright Office acts on his application and he is able to seek redress from the courts. *Nimmer* thoroughly addresses this gap between application and registration and how courts have confronted it.¹⁷¹ Some courts have sought to mitigate the effects of the gap by allowing plaintiffs who filed suit while their applications were pending in the Copyright Office to amend their complaint after the Copyright Office completed registration.¹⁷² Of course, allowing plaintiffs in this situation to amend their complaint after registration does not alleviate the owner's inability to prevent infringement where the application remains pending beyond the time when a court would otherwise be ready to handle the case; the owner must continue to wait on the Copyright Office to act before the court will provide relief. Another registration approach court explained that an applicant who is unable to sue because his application is pending will ultimately be made whole through damages recovered in the suit.¹⁷³ Because a plaintiff's registration is backdated to the date when the Copyright Office received his application, the court contended that the damages sufficiently compensate the plaintiff, albeit after some delay.¹⁷⁴ However, *Nimmer* points out that although the backdating

165. See *supra* Part III (discussing the Seventh and Eighth Circuits).

166. See *supra* text accompanying notes 93-98 (discussing policy rationales supporting the application approach).

167. See generally *supra* Part II.B (discussing application approach arguments).

168. See 17 U.S.C. § 411(a) (2006 & Supp. 2009).

169. See, e.g., *Apple Barrel Prods., Inc. v. Beard*, 730 F.2d 384, 386-87 (5th Cir. 1984).

170. See *supra* note 35 and accompanying text.

171. NIMMER & NIMMER, *supra* note 91, § 7.16(B)(1)(a)(i).

172. *Id.*

173. *Mays & Assocs., Inc. v. Euler*, 370 F. Supp. 2d 362, 369-70 (D. Md. 2005).

174. *Id.*; see 17 U.S.C. § 410(d) (2006) ("The effective date of a copyright registration is the day on which an application, deposit, and fee, which are later determined by the Register of

provision alleviates much of the concern over this “legal limbo” while the plaintiff is awaiting action by the Copyright Office, the plaintiff is still not made completely whole:

[G]iven that belated registration will not allow recovery of statutory damages for infringement that commenced before its effective date, and given further a three-year statute of limitations for recovery of all damages caused by copyright infringement, the [registration] approach may indeed occasion complete inability to recover damages—especially in its hyper-technical instantiation.¹⁷⁵

The harm to plaintiffs caused by making them endure this limbo period before being able to protect their copyrights would only be worthwhile if there were some significant policy justification in favor of delaying institution of infringement suits until the Copyright Office acts. Registration approach advocates have argued that such justification exists because a decision by the Copyright Office to deny registration should be given deference.¹⁷⁶ Naturally, for courts to defer to the Copyright Office, “there must be a resolution from the Copyright Office to which to defer.”¹⁷⁷ Forcing plaintiffs to wait through the period during which their applications are pending ensures that courts will have such a resolution to afford deference. Despite the reasonable logic of this policy argument, *Nimmer* points out that the practical realities of litigation render moot any concerns that federal courts would actually be forced to act without a decision from the Copyright Office:

True, if courts following the . . . [application] approach were inexorably forced to evaluate the strength of the copyright application without guidance from the Copyright Office, there might indeed be reason to force plaintiffs to wait before instituting suit. But, in fact, the pace of litigation entails that the Copyright Office will typically have granted or refused registration during its pendency. Therefore . . . the Register typically will not be deprived of her opportunity, in due course, to appear in the litigation, even if the complaint is allowed to be filed in the interim before the application has been accepted or rejected.¹⁷⁸

Nimmer also proposes that courts require plaintiffs to notify the Copyright Office of litigation so that in “the rare case in which compressed timing might prejudice the Register’s right to appear,” the Register of Copyrights “could expedite treatment of the application and, if she denies issuance of a certificate, could appear in the litigation to defend her determination.”¹⁷⁹ Along with specifying

Copyrights or by a court of competent jurisdiction to be acceptable for registration, have all been received in the Copyright Office.”).

175. NIMMER & NIMMER, *supra* note 91, § 7.16(B)(1)(a)(i).

176. *See id.*

177. *Id.*

178. *Id.*

179. *Id.*

that submission of an application suffices to allow a copyright owner to institute an infringement suit, Congress should include a provision in the Copyright Act reflecting this proposal from *Nimmer* that plaintiffs must notify the Copyright Office of their pending litigation. Under such a statutory scheme, the concerns that court will be forced to act without a decision from the Copyright Office would be further alleviated.

At least one commentator has suggested that the Supreme Court should lead the way by granting certiorari and adopting the application approach.¹⁸⁰ However, under this solution the federal courts would still be acting against the Act's plain meaning, as discussed in the previous subsection.¹⁸¹ Therefore, this Note seeks to place the onus not on the Supreme Court, but on Congress, as "[i]t is the province of the legislature to make the laws, and of the courts to enforce them."¹⁸²

CONCLUSION

Although U.S. copyright law recognizes the rights of authors in their works immediately,¹⁸³ the registration requirement under section 411(a) of the Copyright Act prevents authors from being protected by federal courts without significant delay. Widespread adoption of the registration approach would only continue to place meaningless formalities above copyright owners' need for early protection. As a result, application approach courts are inventing an interpretation of "registration" that is simply not supported by the statute itself, and federal courts throughout the country are split over whether to follow this inaccurate but policy-based application approach. Courts are now confused about what section 411(a) actually requires, although the actual meaning of the language is clear. It is now time for Congress to step in and amend the statute in recognition that only harm is done by requiring authors to wait on the Copyright Office before being able to protect their copyrights.

180. Hogan, *supra* note 16, at 845.

181. See *supra* Part IV.A.

182. *Barrett v. State*, 229 U.S. 26, 30 (1913).

183. 17 U.S.C. § 102(a) (2006).

JUSTICE OR MENTAL HEALTH . . . SHOULD LITIGANTS HAVE TO CHOOSE? MENTAL HEALTH AS A REASON TO PROCEED ANONYMOUSLY

SARAH ORME*

INTRODUCTION

Suppose you are a plaintiff with a civil claim against another party involving your struggle with bipolar disorder, the details of which will necessarily be revealed in the course of litigation. You have only revealed the fact that you suffer from mental illness to your immediate family, your medical providers, and your attorney, and you fear that making the specifics of your mental illness public may jeopardize your personal and professional relationships. You filed a motion to proceed anonymously, but the court has denied it, so you are left with two options: divulge your highly sensitive mental health information or abandon your claim in order to protect your privacy. This is the predicament faced by civil litigants suffering from mental illness under the current laws for proceeding anonymously.

Rule 10(a) of the Federal Rules of Civil Procedure requires that, among other things, “[e]very pleading . . . must name all the parties.”¹ Despite this rule, some courts recognize that in certain circumstances, it is appropriate to allow a party to proceed anonymously.² Anonymity has been consistently granted to protect “children, rape victims, and other particularly vulnerable parties or witnesses”³ and in cases that involve particularly sensitive and personal matters such as “birth control, abortion, or homosexuality.”⁴ However, courts have not demonstrated the same consistency in granting anonymity when a litigant’s reason for wishing to proceed anonymously involves the litigant’s mental illness.⁵ For purposes of this Note, a litigant’s mental illness is “involved” in litigation when it will be an important aspect of either party’s claims or defenses, or when the details of the

* J.D. Candidate, 2011, Indiana University School of Law—Indianapolis; B.A., 2008, Willamette University, Salem, Oregon. I would like to thank Professor Joel Schumm for his help developing this topic, and Duane Marks and Danielle Tucker for their valuable advice and editing. I would also like to thank my husband Philip for his patience and support throughout the note-writing process.

1. FED. R. CIV. P. 10(a).

2. See generally Francis M. Dougherty, Annotation, *Propriety and Effect of Use of Fictitious Names of Plaintiff in Federal Court*, 97 A.L.R.FED. 369 (1990).

3. Doe v. Blue Cross & Blue Shield United of Wis., 112 F.3d 869, 872 (7th Cir. 1997).

4. Dougherty, *supra* note 2, at § 2a.

5. See, e.g., Doe v. Provident Life & Accident Ins. Co., 176 F.R.D. 464 (E.D. Pa. 1997) (finding pseudonymity justified where very few people knew of plaintiff’s mental illness and he feared being stigmatized); Doe v. Gallinot, 486 F. Supp. 983 (C.D. Cal. 1979) (allowing plaintiff to proceed anonymously in case involving his involuntary commitment). But see Doe v. Ind. Black Expo, Inc., 923 F. Supp. 137, 141 (S.D. Ind. 1996) (not allowing plaintiff to proceed anonymously even though his mental health history, including hospitalization, would be part of the litigation).

mental illness will be revealed in discovery and discussed in the course of litigation.

Federal circuit courts vary in how they treat a litigant's request to proceed anonymously for reasons associated with mental illness. Most courts rely on some form of multi-factor test that weighs the litigant's privacy interest against the public interest in knowing who is using the court.⁶ Generally, under such multi-factor tests, the presumption favors public hearings.⁷ However, the application of various multi-factor tests and the uneven weight given to the factors by different courts has resulted in varying outcomes on motions for anonymity in litigation involving mental illness.⁸

The dispositive issue in cases involving a plaintiff who wishes to proceed anonymously (because her private mental health information will be revealed in the course of litigation) is generally whether the case presents an exceptional circumstance. The court in these cases considers whether the plaintiff has demonstrated an "exceptional" circumstance wherein her "substantial privacy right . . . outweighs the 'customary and constitutionally-embedded presumption of openness in judicial proceedings.'"⁹ Despite this established method of analysis, there are no standards for determining what makes a case "exceptional" or which privacy interests are "substantial" enough to outweigh the presumption of openness.¹⁰ Although the terms "exceptional circumstance" and "substantial privacy interest" are not clearly defined, there are both public and private interests that favor a more liberal and clearly defined rule on this issue.¹¹ The public has an interest in protecting the privacy of litigants so that plaintiffs are not deterred from pursuing otherwise valid claims.¹² Private parties—the litigants themselves—have an interest in proceeding anonymously in order to protect the confidentiality of their health information.¹³ Additionally, mental illness is still stigmatized in society, and revealing this sensitive information could have negative effects on a plaintiff's social and professional life, as well as on her continued mental well-being.¹⁴

Part I of this Note provides a brief explanation of mental illness—the

6. See *EW v. N.Y. Blood Ctr.*, 213 F.R.D. 108, 110-11 (E.D.N.Y. 2003).

7. See, e.g., *Doe v. Frank*, 951 F.2d 320, 323-24 (11th Cir. 1992); *Doe v. Stegall*, 653 F.2d 180, 186 (5th Cir. Aug. 1981).

8. See, e.g., *Does I-IV v. City of Indianapolis*, 1:06-cv-865-RLY-WTL, 2006 U.S. Dist. LEXIS 54877, at *7-8 (S.D. Ind. Aug. 7, 2006); *N.Y. Blood Ctr.*, 213 F.R.D. at 112-13.

9. *Frank*, 951 F.2d at 323 (citation omitted).

10. See, e.g., *Anon. v. Legal Servs. Corp. of P.R.*, 932 F. Supp. 49, 51 (D.P.R. 1996) (allowing plaintiff to proceed anonymously in case involving a treatable mental illness). But see *Ind. Black Expo, Inc.*, 923 F. Supp. at 141 (denying plaintiff's petition to proceed anonymously even though his mental health history would be part of the litigation).

11. *Doe v. Provident Life & Accident Ins. Co.*, 176 F.R.D. 464, 467 (E.D. Pa. 1997).

12. *Id.*

13. *Id.*

14. Bethany A. Teachman et al., *Implicit and Explicit Stigma of Mental Illness in Diagnosed and Healthy Samples*, 25 J. SOC. & CLINICAL PSYCHOL. 75, 77 (2006).

individuals affected, how they are affected, and the types and effectiveness of treatment. This section also briefly explains statutory treatment of health information as well as the general history of proceeding anonymously and courts' treatment of the practice. Part II examines how different jurisdictions have treated mental illness as a cause for proceeding anonymously. Part III discusses necessary considerations for a general rule regarding mental illness as a cause for proceeding anonymously in federal court and proposes a new general rule. Finally, Part IV advocates for a rule to be used across jurisdictions—one that incorporates the strengths of the existing rules and advocates for a generally more tolerant approach to the unique interests at stake in cases involving mental health information.

I. BACKGROUND

A. Mental Illness

To completely understand what is at stake in the issue at hand, it is necessary to be aware of the pervasiveness of mental illness in society. Mental illnesses “are medical conditions that disrupt a person’s thinking, feeling, mood, ability to relate to others and daily functioning . . . [and] often result in a diminished capacity for coping with the ordinary demands of life.”¹⁵ One in seventeen Americans lives with a serious mental illness.¹⁶ However, in a given year, as many as one in four American adults will experience mental illness.¹⁷ People of any race, culture, or income may be affected by mental illness,¹⁸ although manifestations vary based on these factors.¹⁹ Additionally, people of all ages are susceptible to mental health disorders, “but the young and the old are especially vulnerable.”²⁰ In fact, mental illness most often strikes during adolescence or young adulthood.²¹

Mental illnesses fall into two general categories: anxiety disorders and mood disorders.²² Examples of anxiety disorders include obsessive-compulsive disorder and post-traumatic stress disorder; examples of mood disorders include bipolar

15. *What Is Mental Illness: Mental Illness Facts*, NAT’L ALLIANCE ON MENTAL ILLNESS, http://www.nami.org/Content/NavigationMenu/Inform_Yourself/About_Mental_Illness/About_Mental_Illness.htm (last visited Feb. 5, 2011).

16. *Id.* “Serious mental illness” includes major depression, schizophrenia, bipolar disorder, obsessive-compulsive disorder (OCD), panic disorder, post-traumatic stress disorder (PTSD), and borderline personality disorder. *Id.*

17. *Id.*

18. *Id.*

19. OFFICE OF THE SURGEON GEN., MENTAL HEALTH: A REPORT OF THE SURGEON GENERAL ch. 2, available at http://www.surgeongeneral.gov/library/mentalhealth/chapter2/sec2_1.html (last visited Feb. 5, 2011).

20. *What Is Mental Illness: Mental Illness Facts*, *supra* note 15.

21. *See id.*

22. OFFICE OF THE SURGEON GEN., *supra* note 19.

disease and schizophrenia.²³ Generally, mental illnesses manifest as “clusters of symptoms and signs” that impair a person’s ability to function and are often triggered by a combination of biological, psychological, and socio-cultural factors.²⁴ The presence of these risk factors, which come together in a complex chain of causation (often triggered by a stressful life event), increases the probability that a person will develop a disorder.²⁵ Common manifestations of mental illness include phobias, panic attacks, hallucinations, delusions, depression, and mania.²⁶

Although mental illness is prevalent in our society, the treatments available for mental illness have come a long way in recent decades. Mental illness was once viewed as a lifelong deterioration with little hope for improvement.²⁷ Today, however, new medications treat even severe mental illness and allow most afflicted individuals some relief from their symptoms.²⁸ In fact, most people suffering from mental illness—seventy to ninety percent—can experience a reduction of symptoms and an improved quality of life with a combination of medication, therapy, and support.²⁹ Available support options for those suffering from mental illness include self-help, mental health consumer, and advocacy groups.³⁰ Unfortunately, society’s understanding of mental illness does not always reflect the same advancements.

The stigmas that have long been associated with mental illness remain staunchly in place today. In fact, at least one commentator argues that “society’s continued stigmatizing response to mental illness makes it one of the most marginalized conditions in modern Western societies.”³¹ The stigmatization that individuals with mental illnesses experience often results in “decreased life opportunities and a loss of independent functioning over and above the impairments related to mental disorders themselves.”³² There is still significant evidence that “the label ‘mentally ill’ makes it more difficult to obtain work and housing, and to gain acceptance from peers and co-workers, regardless of the individual’s behavior.”³³ Thus, despite the improved understanding of the causes, manifestations, and treatments for mental illness in the scientific community, being labeled “mentally ill” continues to have significant negative connotations and consequences in professional and social life.

23. *Id.*

24. *Id.*

25. *Id.*

26. *Id.*

27. *See id.*

28. *What Is Mental Illness: Mental Illness Facts*, *supra* note 15.

29. *Id.*

30. OFFICE OF THE SURGEON GEN., *supra* note 19.

31. Teachman, *supra* note 14, at 92.

32. Stephen P. Ninshaw & Andrea Stier, *Stigma as Related to Mental Disorders*, 4 ANN. REV. CLINICAL PSYCHOL. 367, 367 (2008).

33. Teachman, *supra* note 14, at 77.

B. Treatment of Health Information Generally

In order to fully appreciate the treatment of motions to proceed anonymously for reasons related to mental illness in federal courts, it is important to understand how other areas of the law treat health information. In the area of health information regulations, health records are generally considered to be confidential.³⁴ The federal government and many state governments have passed legislation to ensure this confidentiality.³⁵ One of the most well-known and widely applicable health information privacy statutes is the Health Insurance Portability and Accountability Act (HIPAA), enacted by the Congress in 1996.³⁶ The HIPAA Privacy Rule “gives . . . [individuals] rights over . . . [their] health information and sets rules and limits on who can look at and receive . . . [this] health information.”³⁷

HIPAA requires certain health care entities—including health plans, most health care providers, and health care clearinghouses—to protect health information by putting safeguards in place, reasonably limiting disclosure of health information to the minimum necessary to accomplish the purpose of the disclosure, and limiting who can view and access personal health information.³⁸ If an entity covered by HIPAA is required to disclose protected health information for litigation purposes, the entity “must make reasonable efforts to limit the protected health information disclosed to the minimum necessary” for the purpose of the disclosure; “this could involve de-identifying the information or stripping direct identifiers from the information to protect the privacy of individuals.”³⁹ Thus, HIPAA provides important protections and rights to consumers regarding the privacy of their health care information.

In addition to the federal HIPAA regulations, states can pass additional and more stringent statutes regarding the privacy of health information that will not be preempted by HIPAA unless they are contrary to the objectives of the federal statute.⁴⁰ To date, a number of states have passed such additional health information privacy statutes that are “more protective of the records of mental patients than they have been of medical records generally.”⁴¹ The Indiana

34. 62A AM. JUR. 2D *Privacy* § 183 (2010).

35. See, e.g., Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 101 Stat. 1936; see also IND. CODE § 16-39-3-10 (2010).

36. See *Health Information Privacy*, U.S. DEP’T OF HEALTH & HUMAN SERVS., <http://www.hhs.gov/ocr/privacy/index.html> (last visited Feb. 5, 2011).

37. *Health Information Privacy for Consumers*, U.S. DEP’T OF HEALTH & HUMAN SERVS., <http://www.hhs.gov/ocr/privacy/hipaa/understanding/consumers/index.html> (last visited Feb. 5, 2011).

38. See *id.*

39. *Health Information Privacy: Frequently Asked Questions*, U.S. DEP’T OF HEALTH & HUMAN SERVS., http://www.hhs.gov/ocr/privacy/hipaa/faq/judicial_and_administrative_proceedings/705.html (last visited Feb. 5, 2011).

40. See 39A C.J.S. *Health & Env’t* § 4 (2009).

41. 56 C.J.S. *Mental Health* § 17 (2009).

legislature, for example, passed Indiana Code section 16-39-3-10 regarding the confidentiality of mental health information that reflects this common practice. Indiana Code section 16-39-3-10 states, “If a patient’s mental health record or testimony related to a patient’s mental health is offered or admitted into evidence in a legal proceeding, the court shall maintain the record or transcript of the testimony as a confidential court record.”⁴² The Indiana Code addresses the release of mental health records in investigations and legal proceedings, recognizing the privacy interest at stake and the need for confidentiality as well as the concern that disclosure of these records may have negative repercussions on a patient’s mental health rehabilitation.⁴³

These federal and state statutes clearly recognize that health information—especially information relating to mental health—is considered an important individual privacy interest that should be guarded by the government. Accordingly, the government generally affords health information special protection and takes steps to keep such information confidential.

C. Proceeding Anonymously

The ability to proceed anonymously in a trial “provides the plaintiff in many cases with the only means to pursue important substantive rights” by allowing plaintiffs to protect their privacy while pursuing meritorious claims they might otherwise give up.⁴⁴ The practice of proceeding anonymously using the pseudonym “John Doe” began in the seventeenth century, but it was not until the 1960s, when the Supreme Court recognized the right to privacy under the Constitution, that plaintiffs began to use the pseudonym to hide their identities.⁴⁵ Jurisdictions allowing plaintiffs to proceed anonymously cite reasons associated with recognized privacy rights under the Constitution.⁴⁶ In general, these jurisdictions first recognize that plaintiffs may want to remain anonymous due to fears of “public stigma, personal safety, and economic retribution.”⁴⁷ Second, they acknowledge that the information plaintiffs will have to disclose may be “too intimate to disclose publicly.”⁴⁸ Third, and most importantly, these jurisdictions appreciate that plaintiffs may forgo their meritorious claims because they fear revealing their private information.⁴⁹

Jurisdictions that do not allow plaintiffs to proceed anonymously generally

42. IND. CODE § 16-39-3-10 (2010).

43. See *id.* § 16-39-3-9.

44. Carol M. Rice, *Meet John Doe: It Is Time for Federal Civil Procedure to Recognize John Doe Parties*, 57 U. PITTS. L. REV. 883, 886 (1996).

45. See *id.* at 889-94.

46. See *id.* at 908.

47. *Id.*

48. *Id.*

49. Jayne S. Ressler, *Privacy, Plaintiffs, and Pseudonyms: The Anonymous Doe Plaintiff in the Information Age*, 53 U. KAN. L. REV. 195, 219 (2004).

rely on reasoning based on ideas of fairness to the public and the defendant.⁵⁰ These jurisdictions often cite the argument that the presumption in favor of open judicial proceedings and the public's right to be informed regarding the proceedings, found in the First Amendment, are the primary interests that conflict with a plaintiff's ability to proceed anonymously.⁵¹ Second, these jurisdictions argue that the use of pseudonyms may prejudice the defendant by making it difficult to perform discovery and form defenses and counterclaims.⁵² Third, they rely on the rule that is cited most often in opposition to a plaintiff's motion to proceed anonymously—Federal Rule of Civil Procedure Rule 10(a) (“FRCP 10(a)”).⁵³

II. ANONYMITY IN FEDERAL COURTS

A. In General

The test used by many federal courts to determine when an exception to FRCP 10(a) is appropriate and whether a plaintiff should be allowed to proceed anonymously has evolved through case law over the last three decades.⁵⁴ In *Doe v. Stegall*,⁵⁵ the Fifth Circuit identified three characteristics common to “those exceptional cases in which the need for party anonymity overwhelms the presumption of disclosure mandated by procedural custom.” The court listed the factors as follows: “(1) plaintiffs seeking anonymity were suing to challenge governmental activity; (2) prosecution of the suit compelled plaintiffs to disclose information ‘of the utmost intimacy,’ and (3) plaintiffs were compelled to admit their intention to engage in illegal conduct, thereby risking criminal prosecution.”⁵⁶ The court did not intend for these characteristics to form a “hard and fast formula;” rather, it hoped for a balancing of considerations.⁵⁷ Later, the Fifth Circuit elaborated on its *Stegall* test in *Doe v. Frank*, clarifying that the “ultimate test . . . is whether the plaintiff has a substantial privacy right which outweighs the ‘customary and constitutionally-embedded presumption of

50. See *id.* at 212; see also Mark Albert Mesler II, *Civil Procedure—Doe v. Frank: Determining the Circumstances Under Which a Plaintiff May Proceed Under a Fictitious Name*, 23 MEM. ST. U. L. REV. 881, 882 (1993).

51. See Ressler, *supra* note 49, at 212.

52. See Mesler, *supra* note 50, at 882.

53. See FED. R. CIV. P. 10(a), which states: “Every pleading must have a caption with the court’s name, a title, a file number, and a Rule 7(a) designation. The title of the complaint must name all the parties; the title of other pleadings, after naming the first party on each side, may refer generally to other parties.”

54. See generally *Doe v. Frank*, 951 F.2d 320, 323-24 (11th Cir. 1992); *Doe v. Stegall*, 653 F.2d 180 (5th Cir. 1981); *EW v. N.Y. Blood Ctr.*, 213 F.R.D. 108 (E.D.N.Y. 2003).

55. *Stegall*, 653 F.2d at 180.

56. *Id.* at 185.

57. See *id.* at 186.

openness in judicial proceedings.”⁵⁸ Further, the court concluded that the three factors laid out in *Stegall* were factors for a court to consider in making its determination.⁵⁹

In *EW v. N.Y. Blood Center*,⁶⁰ the court pulled from *Stegall*, *Frank*, and other cases across jurisdictions to create a six-factor test to determine “whether a plaintiff’s privacy right outweighs the public interest in open proceedings and any possible prejudice to the defendant.”⁶¹ The six-factor test involved the determinations of:

- (1) whether the plaintiff is challenging governmental activity or an individual’s actions;
- (2) whether the plaintiff’s action requires disclosure of information of the utmost intimacy;
- (3) whether the action requires disclosure of the plaintiff’s intention to engage in illegal conduct;
- (4) whether identification would put the plaintiff at risk of suffering physical or mental injury;
- (5) whether the defendant would be prejudiced by allowing the plaintiff to proceed anonymously; and
- (6) the public interest in guaranteeing open access to proceedings without denying litigants access to the judicial system.⁶²

Most recently, in *Does I-IV v. City of Indianapolis*,⁶³ the court adopted the “ultimate test” from *Doe v. Frank*—weighing the plaintiff’s substantial privacy right against the presumption of open court proceedings—and adopted the six-factor test from *New York Blood Center* to assess the balance of the two opposing interests.⁶⁴

The series of tests used by courts over the past three decades and the policy underlying each of them—allowing plaintiffs to proceed anonymously when their privacy interests outweigh public interests—has worked fairly well to protect certain groups of plaintiffs. For example, women seeking abortions, homosexuals,⁶⁵ children, and rape victims⁶⁶ are generally able to proceed anonymously under the various tests discussed above. However, a large block of the population is not consistently protected under these analyses and should be.

Plaintiffs suffering from mental illness—or who have mental health issues in their past which litigation will reveal—make up one group of individuals this author believes should be afforded the protection of proceeding anonymously.

58. *Frank*, 951 F.2d at 323 (quoting *Stegall*, 653 F.2d at 186).

59. *Id.*

60. 213 F.R.D. 108 (E.D.N.Y. 2003).

61. *Id.* at 111 (citations omitted).

62. *Id.* (internal citations omitted).

63. 1:06-cv-865-RLY-WTL, 2006 U.S. Dist. LEXIS 54877 (S.D. Ind. Aug. 7, 2006).

64. *Id.* at *4.

65. See, e.g., Dougherty, *supra* note 2, § 29.

66. See *Doe v. Blue Cross & Blue Shield United of Wis.*, 112 F.3d 869, 872 (7th Cir. 1997).

Plaintiffs with mental health issues have a unique and significant interest in protecting this highly sensitive information, as there are still stigmas in our society associated with mental illness.⁶⁷ Additionally, having to make this private information public may be a stressful life event that could re-trigger a plaintiff's mental illness or negatively affect her rehabilitative process.⁶⁸ These privacy implications are even greater today given the increased accessibility of information, including judicial opinions, through the Internet.⁶⁹

As it relates to this issue, the law should be clarified so that all groups that need the protection of anonymity in the courtroom are entitled to it. The question, however, is how best to achieve this protection. The common law rule that has evolved over the past three decades needs to be reevaluated to take the unique issues associated with mental illness into consideration. But should courts or Congress determine when an exception to FRCP 10(a) is appropriate? Courts are arguably not the most suitable forum for creating a general rule or policy regarding mental illness as a cause to proceed anonymously. Thus far, common law has produced a jumble of rules that often lead to inconsistent results. Additionally, courts may not be the appropriate body to take on the task of challenging the well-entrenched stigmas in our society associated with mental illness. Congress, however, also may not be the ideal forum for determining which situations justify allowing a plaintiff to proceed anonymously, as many members of Congress may not have the legal experience necessary to know what kinds of procedural rules are realistic in a courtroom setting.

The Judicial Conference of the United States ("Judicial Conference"), which was created by Congress in 1922 to make policy for the U.S. courts, provides a perfect forum for creating a new rule for determining whether a plaintiff should be allowed to proceed anonymously when she must reveal mental health information in litigation. The Judicial Conference is the ideal forum because it combines the strengths that courts and Congress each have in creating policy.⁷⁰ Moreover, it addresses and advises courts on a variety of subjects including rules of practice and procedure.⁷¹ Like Congress, the Judicial Conference may hold hearings and take sworn testimony to inform its policymaking.⁷² The opportunity for the Judicial Conference to hear from advocacy groups and experts on mental

67. See discussion *supra* Part I.A.

68. See *id.*; see also IND. CODE § 16-39-3-9 (2010) (requiring judges to limit release of patients' mental health information to protect the rehabilitative process).

69. See Joel M. Schumm, *No Names, Please: The Virtual Victimization of Children, Crime Victims, the Mentally Ill, and Others in Appellate Court Opinions*, 42 GA. L. REV. 471, 475-76 (2008).

70. See *Judicial Conference of the United States*, U.S. COURTS, <http://www.uscourts.gov/judconf.html> (last visited Feb. 5, 2011). The Conference, as originally created in 1922, was called the Conference of Senior Circuit Judges. Congress enacted 28 U.S.C. § 331 in 1948, which changed the name to the Judicial Conference of the United States. *Id.*

71. See *Judicial Conference of the United States: Organization*, U.S. COURTS, <http://www.uscourts.gov/FederalCourts/JudicialConference/Organization.aspx> (last visited Feb. 5, 2011).

72. 28 U.S.C. § 331 (2006).

illness as it pertains to litigation would be beneficial in formulating a new procedural rule regarding when the release of mental health information justifies allowing a plaintiff to proceed anonymously. Additionally, because the Judicial Conference is comprised of the Chief Justice of the United States Supreme Court as well as circuit and district court judges, the people who will eventually apply this new rule will also be instrumental in making it.⁷³ Thus, the Judicial Conference is in a position to consider the unique interests of plaintiffs who are concerned about the privacy of their mental health information—as well as the realities of litigation—as it formulates a new rule of procedure to address plaintiff anonymity in situations involving the plaintiff's mental health information.

B. Mental Illness as Cause for Proceeding Anonymously

1. Federal Circuits That Allow Mental Illness as a Cause to Proceed Anonymously.—District courts in four federal circuits—the First, Second, Third, and Eleventh—have consistently allowed plaintiffs to proceed anonymously due to mental health implications.⁷⁴ In the Eastern District of New York case *Doe No. 2 v. Kolko*,⁷⁵ the plaintiff moved to proceed anonymously, claiming that he suffered from post-traumatic stress disorder, bipolar disorder, and depression, and that he would suffer psychological harm if his identity was revealed in litigation.⁷⁶ The court applied a five-factor test to determine whether the plaintiff's need for anonymity outweighed the “prejudice to the opposing party and the public's interest in knowing the party's identity.”⁷⁷ Ultimately, the court found that the intimate nature of the complaint, which involved sexual abuse along with the plaintiff's fragile psychological condition, “established special circumstances to warrant authorization to proceed anonymously.”⁷⁸

This trend in district courts in the Second Circuit allowing plaintiffs to proceed anonymously in cases involving their mental health was also demonstrated in two earlier cases. In *Doe v. New York University*,⁷⁹ the plaintiff claimed New York University had discriminated against her on the basis of her mental illness despite the fact that she had undergone psychiatric treatment and had “regained sufficient emotional stability to return to school.”⁸⁰ In *Doe v.*

73. *Judicial Conference of the United States: Membership*, U.S. COURTS, <http://www.uscourts.gov/FederalCourts/JudicialConference/Membership.aspx> (last visited Feb. 5, 2011).

74. See, e.g., L.C. v. Olmstead, 138 F.3d 893 (11th Cir. 1998); Doe No. 2 v. Kolko, 242 F.R.D. 193 (E.D.N.Y. 2006) (Second Circuit); Doe v. Provident Life & Accident Ins. Co., 176 F.R.D. 464 (E.D. Pa. 1997) (Third Circuit); Anon. v. Legal Servs. Corp. of P.R., 932 F. Supp. 49 (D.P.R. 1996) (First Circuit).

75. 242 F.R.D. 193 (E.D.N.Y. 2006).

76. See *id.* at 194-95.

77. *Id.* at 195 (citation omitted).

78. *Id.* at 196, 198.

79. 442 F. Supp. 522 (S.D.N.Y. 1978).

80. *Id.* at 522.

Harris,⁸¹ the plaintiff sought review of a decision of the Department of Health and Human Services that denied his application for disability benefits despite the fact that he had been committed to mental institutions several times and had been diagnosed with schizophrenia.⁸² In both of these cases, the plaintiffs were allowed to proceed anonymously. However, neither opinion provided any explanation for the court's decision to grant anonymity.⁸³

In the Third Circuit, district courts have also demonstrated a trend of granting motions to proceed anonymously in mental health cases. In *Doe v. Provident Life & Accident Insurance Co.*,⁸⁴ the District Court for the Eastern District of Pennsylvania found that the plaintiff had sufficiently justified his pseudonymity because he had only revealed his mental illness to his immediate family.⁸⁵ Furthermore, the plaintiff feared being stigmatized by his community, which could include damage to his professional life, if his illness became public.⁸⁶ The standard applied by the district court stated that the "public's . . . right of access should prevail unless . . . [the party requesting pseudonymity] demonstrates . . . that his interests in privacy or security justify pseudonymity."⁸⁷ The court applied a multi-factor test to weigh the public and private interests involved.⁸⁸ In granting the plaintiff's motion to proceed anonymously, the court gave great weight to the possibility that litigants with mental illnesses could be stigmatized and that fear of stigmatization may deter people whose mental illness will be an important aspect of litigation from pursuing claims.⁸⁹ Subsequently, district courts in the Third Circuit have allowed plaintiffs in many other cases—all involving details of a plaintiff's mental illness—to proceed using pseudonyms in order to protect the plaintiff's identity.⁹⁰

Likewise, courts in the Eleventh and First Circuits have allowed plaintiffs to proceed anonymously when details of their mental illnesses were part of the litigation.⁹¹ In *L.C. by Zimring v. Olmstead*, an Eleventh Circuit Court of Appeals case involving the plaintiff's involuntary confinement for mental illness, the court allowed the plaintiff to proceed anonymously.⁹² The court did not discuss this decision in any detail, as it was deferring to the decision made by the district court to allow the plaintiff to proceed using a pseudonym in order to protect her

81. 495 F. Supp. 1161 (S.D.N.Y. 1980).

82. See *id.* at 1163-66.

83. See *id.* at 1161; *New York Univ.*, 442 F. Supp. at 522.

84. 176 F.R.D. 464 (E.D. Pa. 1997).

85. See *id.* at 468.

86. *Id.*

87. *Id.* at 467 (citation omitted).

88. See *id.* at 468-69.

89. See *id.* at 468.

90. See, e.g., *PAS v. Travelers Ins. Co.*, 7 F.3d 349 (3d Cir. 1993), *Doe v. Colautti*, 592 F.2d 704 (3d Cir. 1979), *Doe v. Hartford Life & Accident Ins. Co.*, 237 F.R.D. 545 (D.N.J. 2006).

91. See *L.C. v. Olmstead*, 138 F.3d 893 (11th Cir. 1998); *Anon. v. Legal Servs. Corp. of P.R.*, 932 F. Supp. 49 (D.P.R. 1996).

92. *Olmstead*, 138 F.3d at 895 n.1.

identity.⁹³ In *Anonymous v. Legal Services Corp.*, the court allowed the plaintiff to proceed anonymously because her allegations involved a treatable mental illness, which was a privacy interest sufficient to justify pseudonymity.⁹⁴

The circuits that allow mental illness as a cause to proceed anonymously base their decisions on two primary policy considerations.⁹⁵ First, federal courts traditionally permit parties to proceed anonymously when the parties have a strong privacy interest in doing so.⁹⁶ For example, a mental illness that could harm one's profession would be considered one such privacy interest.⁹⁷ Second, these courts recognize that cases involving mental illness require special consideration because of the stigma surrounding mental illness.⁹⁸ This consideration can be further divided into the interests of litigants and the interests of the public, both of which—contrary to the argument used in circuits that have disfavored plaintiff anonymity—support allowing plaintiffs to proceed anonymously. One interest is that litigants with mental illness have a strong interest in protecting their privacy while retaining the ability to vindicate their rights through litigation.⁹⁹ The second is that the public has a strong interest in preventing the stigmatization of litigants and protecting plaintiffs' privacy so that plaintiffs are not discouraged from pursuing their claims.¹⁰⁰

2. Federal Circuits That Do Not Allow Mental Illness as a Cause to Proceed Anonymously.—Two federal circuits—the Seventh and the Tenth—have consistently denied plaintiffs permission to proceed anonymously in cases involving a plaintiff's mental health. In *Doe v. Blue Cross & Blue Shield United*,¹⁰¹ the Seventh Circuit Court of Appeals found that the plaintiff should not have been permitted to proceed anonymously because the use of fictitious names was “disfavored” and the plaintiff's obsessive-compulsive disorder was not so uncommon or humiliating that it should have been an automatic ground for proceeding anonymously.¹⁰² Similarly, in *Doe v. Indiana Black Expo*,¹⁰³ a district court in the Seventh Circuit held that although litigation of the plaintiff's claims would require detailed consideration of his history of mental health hospitalization and substance abuse, the plaintiff's privacy interest was not sufficient to overcome the strong presumption in favor of open court

93. See *id.*

94. *Legal Servs. Corp.*, 932 F. Supp. at 51.

95. See generally *Doe v. Provident Life & Accident Ins. Co.*, 176 F.R.D. 464 (E.D. Pa. 1997); *Legal Servs. Corp.*, 932 F. Supp. at 49.

96. See, e.g., *Provident Life & Accident Ins. Co.*, 176 F.R.D. at 468; *Legal Servs. Corp.*, 932 F. Supp. at 50.

97. *Legal Servs. Corp.*, 932 F. Supp. at 50.

98. See *Provident Life & Accident Ins. Co.*, 176 F.R.D. at 468.

99. *Id.*

100. *Id.*

101. 112 F.3d 869 (7th Cir. 1997).

102. See *id.* at 872.

103. 923 F. Supp. 137 (S.D. Ind. 1996).

proceedings.¹⁰⁴ Similarly, in the Tenth Circuit case *Raiser v. Brigham Young University*,¹⁰⁵ the court of appeals found that the plaintiff's mention of a history of mental illness in his motion to proceed anonymously was not sufficient to create an "exceptional" case justifying pseudonymity.¹⁰⁶

Both the Seventh and the Tenth Circuits, which disfavor allowing plaintiffs to proceed anonymously,¹⁰⁷ base their reasoning on four primary policy considerations. First, these circuits believe that making common mental disorders an automatic ground for proceeding anonymously would propagate the stigma surrounding mental illness and "the view that mental illness is shameful."¹⁰⁸ Second, they argue that allowing plaintiffs to proceed anonymously would hamper defendants' ability to defend themselves.¹⁰⁹ Third, there is a common belief that when a plaintiff's claims make allegations about the defendant's integrity and reputation, the plaintiff should defend those claims publicly rather than "use his privacy interests as a shelter from which he can safely hurl these accusations."¹¹⁰ Finally, these jurisdictions rely on the longstanding argument that "[l]awsuits are public events."¹¹¹

The decisions made in the Seventh and Tenth Circuits have implications far beyond the outcomes of each individual case. The result of the laws used in these circuits is that plaintiffs with mental illnesses they wish to keep private are left with two options. One option is for the plaintiff to divulge the details of her mental illness, risking stigmatization by the community, social and professional consequences, and even her own mental health. The second option is for the plaintiff to abandon her claim in order to protect her highly sensitive and personal health information, thereby forgoing meritorious claims and rights. These options leave plaintiffs with a choice between two extremes: a day in court or privacy. The pervasive stigmas associated with mental illness, the inconsistent treatment of mental illness as a cause to proceed anonymously among jurisdictions, and the problematic options that many plaintiffs with mental illness face indicate that this issue must be addressed with more careful consideration of the unique issues involved.

C. A Current Controversy: Doe v. The Individual Members of the Indiana State Board of Law Examiners

In July 2009, plaintiff "Jane Doe" filed a claim against the Indiana State Board of Law Examiners in the United States District Court for the Southern

104. See *id.* at 141-42.

105. 127 F. App'x 409 (10th Cir. 2005).

106. See *id.* at 410-11.

107. See, e.g., *Femedeer v. Haun*, 227 F.3d 1244 (10th Cir. 2000); *Blue Cross & Blue Shield United*, 112 F.3d at 869; *Ind. Black Expo, Inc.*, 923 F. Supp. at 137.

108. *Blue Cross & Blue Shield United*, 112 F.3d at 872.

109. *Ind. Black Expo, Inc.*, 923 F. Supp. at 141.

110. *Id.* at 142.

111. *Femedeer*, 227 F.3d at 1246 (quoting *Doe v. Frank*, 951 F.2d 320, 324 (11th Cir. 1992)).

District of Indiana.¹¹² The plaintiff was an attorney admitted to the Illinois bar who wanted to sit for the Indiana bar exam.¹¹³ She claimed that the Indiana State Board of Law Examiners' extensive questioning and requirements for applicants with psychological disorders violated the Americans with Disabilities Act.¹¹⁴ Having previously been diagnosed with an anxiety disorder and post-traumatic stress disorder, she sought to proceed anonymously because she feared she would suffer injury and stigmatization if her mental health history became public knowledge.¹¹⁵ The issue in this case was whether exceptional circumstances existed such that the harm to the plaintiff in revealing her mental health history exceeded the likely harm—namely, the harm that departing from the presumption that parties' identities are public information would cause.¹¹⁶

On August 8, 2009, U.S. Magistrate Judge Tim Baker denied the plaintiff's motion to proceed anonymously and seal the affidavit containing her actual name.¹¹⁷ The court applied the six-factor test used in *Does I-IV v. City of Indianapolis* and reasoned that the plaintiff's anxiety disorder and post-traumatic stress disorder were common and, consequently, should not be considered shameful.¹¹⁸ The court went on to explain that litigants often have to disclose sensitive information, noting that this is the reality of the "sometimes gritty world of litigation."¹¹⁹ Furthermore, Judge Baker relied on the fact that the Seventh Circuit has "repeatedly expressed disapproval of anonymous litigants."¹²⁰

The plaintiff subsequently filed an objection to the judge's order and a motion to review and reverse his decision. On January 4, 2010, U.S. District Judge William Lawrence denied her objection.¹²¹ Judge Lawrence explained that the standard of review for discovery-related decisions made by the magistrate judge was extremely deferential.¹²² He concluded that while "this . . . [was] certainly a close case, the Court . . . [could not] find that the Magistrate Judge's order was clearly erroneous or contrary to law."¹²³ Judge Lawrence gave the plaintiff fourteen days to file an amended complaint identifying herself by name.¹²⁴

When the plaintiff's motion and subsequent objection were denied, she was

112. Doe v. Individual Members of the Ind. State Bd. of Law Exam'rs, No. 1:09-cv-0842-WTL-TAB, 2009 U.S. Dist. LEXIS 69609 (S.D. Ind. Aug. 8, 2009).

113. *Id.* at *1.

114. *Id.*

115. *Id.* at *1, *5.

116. See *id.* at *10-11.

117. *Id.* at *11.

118. *Id.* at *2-5.

119. *Id.* at *6.

120. *Id.* at *10.

121. Doe v. Individual Members of the Ind. State Bd. of Law Exam'rs, No. 1:09-cv-842-WTL-JMS, 2010 U.S. Dist. LEXIS 1001, at *1 (S.D. Ind. Jan. 4, 2010).

122. *Id.* at *1-2.

123. *Id.* at *5.

124. *Id.*

left with the two options discussed above. She could divulge her mental health history, risk stigmatization in both her social and professional life, and possibly jeopardize her own mental health by reversing the recovery steps she had made thus far. Alternatively, she could drop the case, thereby ensuring the privacy of her mental health history but forgoing her claim. Evidently, the plaintiff chose the first of these two options—a subsequent entry on an amended motion for class certification was captioned *Perdue v. The Individual Members of the Indiana State Board of Law Examiners*, and the entry identified the plaintiff as Amanda Perdue.¹²⁵

III. NECESSARY CONSIDERATIONS AND A PROPOSAL FOR A NEW RULE REGARDING MENTAL ILLNESS AS A REASON TO PROCEED ANONYMOUSLY

As illustrated by the cases discussed above, a new rule that directly addresses the issue of mental illness as a cause to proceed anonymously is necessary. Currently, there is little consistency regarding when courts grant anonymity, but on a more basic administrative level, the courts are also inconsistent in how they treat motions to proceed anonymously.¹²⁶ Some courts require that a plaintiff seek leave of the court before submitting a pleading using a pseudonym.¹²⁷ Other courts require the moving party to make a good faith effort to resolve the issue with the opposing party before submitting the motion to the court;¹²⁸ others allow plaintiffs to amend their pleadings to provide their full names.¹²⁹ Some courts simply dismiss the case if the plaintiff does not include her full name.¹³⁰ This procedure for judicial treatment of litigants' motions needs to be consistent across courts and jurisdictions so that litigants and their attorneys know how to proceed when requesting anonymity. Additionally, courts should be deciding this often dispositive issue on the specific facts of each case rather than arbitrarily throwing out cases for failure to comply with an individual judge's preferences. Ideally, all federal courts would adopt a new rule, similar to the one proposed below, to determine whether a plaintiff should be allowed to proceed anonymously. The adoption of a single, new rule by all federal courts would make the administrative and substantive treatment of this issue consistent.

Once the procedure for judicial treatment of motions to proceed anonymously is consistent across jurisdictions, the Judicial Conference needs to create and implement a consistent rule regarding when an exception to FRCP 10(a) is

125. *Perdue v. Individual Members of the Ind. State Bd. of Law Exam'rs*, No. 1:09-cv-842-WTL-JMS, 2010 WL 412028, at *1 (S.D. Ind. Jan. 29, 2010).

126. See Rice, *supra* note 44, at 918.

127. See *id.* at 918-19.

128. E.g., *M.M. v. Zavaras*, 139 F.3d 798, 799-800 (10th Cir. 1998).

129. See, e.g., *Doe v. Individual Members of the Ind. State Bd. of Law Exam'rs*, No. 1:09-cv-0842-WTL-TAB, 2009 U.S. Dist. LEXIS 69609, at *10 (S.D. Ind. Aug. 8, 2009) (denying plaintiff's motion to proceed anonymously but allowing her fourteen days to file an amended complaint identifying herself by name).

130. See Rice, *supra* note 44, at 918-19 & 919 n.123.

appropriate. No satisfactory test currently exists in the federal court system to determine when a plaintiff is allowed to proceed anonymously in civil lawsuits involving the plaintiff's mental health information. However, as discussed above, state and federal governments have passed legislation putting safeguards in place to protect the confidentiality of this information.¹³¹ Although this legislation admittedly addresses different circumstances than those discussed in this Note, the safeguards from these statutes may be helpful in determining what the test for proceeding anonymously should look like. After all, citizens' privacy rights in their health information comprise the central issue in these health information privacy statutes as well as in determining whether a plaintiff should be allowed to proceed anonymously. Accordingly, it makes sense to look at the well-developed statutes already in place to inform the discussion of how to formulate a new test for federal judges to apply.

A. Necessary Considerations for a New Federal Rule

A number of factors must be taken into consideration in a new rule determining when plaintiffs are allowed to proceed anonymously in cases involving their mental health information. For instance, the language commonly used in the existing rules—"exceptional" circumstance and "substantial privacy right"¹³²—is very general and does not take into consideration the unique issues at stake when one's mental health information is involved. Accordingly, this language must be carefully redefined to make it more consistent with the federal and state health information privacy statutes.

Under the current rules, "exceptional" circumstance usually requires that a particular plaintiff's need for anonymity be so substantial that it outweighs the presumption in favor of open court proceedings.¹³³ HIPAA, the Indiana Code, and other laws relating to mental health information suggest that the opposite should be true—an exceptional circumstance should be required to allow an exemption to the confidentiality of medical information.¹³⁴ Admittedly, confidentiality should "yield to the needs of justice, or in the face of a countervailing public interest,"¹³⁵ but these should be situations in which the person wishing to proceed anonymously is a danger to the public or herself.¹³⁶ The public interest in open proceedings, which is often cited as an important factor in these cases, should not be considered an "exceptional" circumstance because allowing a plaintiff to proceed anonymously does not interfere with the

131. See discussion *supra* Part I.B.

132. Doe v. Frank, 951 F.2d 320, 323 (11th Cir. 1992).

133. See *id.*

134. See, e.g., IND. CODE § 16-39-3-10 (2010) (requiring courts to keep mental health records or testimony related to a patient's mental health confidential); *Health Information Privacy: Frequently Asked Questions*, *supra* note 39 (requiring health care entities to make reasonable efforts to protect health information, including de-identifying it).

135. 56 C.J.S. *Mental Health* § 17 (2009) (internal citation omitted).

136. *Id.*

public's ability to stay informed about what is going on in the court.¹³⁷ "Party anonymity does not obstruct the public's view of the issues joined or the court's performance in resolving them. . . . [The] crucial interests served by open trials are not inevitably compromised by allowing a party to proceed anonymously."¹³⁸ Thus, this public interest should not be considered so "exceptional" that it justifies an exemption to the confidentiality of medical records.

The current common law rules are similarly vague about what should be considered in determining whether a "substantial privacy interest" exists in a particular case. The Indiana Code requires judges authorizing the release of mental health information to "limit disclosure for the protection of the patient . . . and the rehabilitative process."¹³⁹ Correspondingly, federal judges should consider the protection of the patient and the patient's rehabilitative process when determining whether a substantial privacy interest exists. In other words, if disclosure of mental health information may reasonably be expected to harm a patient or her mental health rehabilitation process, the court should find that a substantial privacy right exists and should consequently protect that privacy right by allowing the plaintiff to proceed anonymously.

Another factor the Judicial Conference must consider in formulating this new rule is the court's general approach to mental illness. Federal courts should adopt an approach that is more tolerant of mental illness and the unique issues mental illness raises for those who suffer from it. Federal and state statutes governing the privacy of health information, especially when it involves mental health, provide good examples for a more tolerant approach because they favor protecting the confidentiality of this sensitive information.¹⁴⁰ This need for a more tolerant approach is particularly relevant to the conflict between open judicial proceedings and the requirement that a plaintiff divulge her mental health information without anonymity. Federal courts should follow the lead of Indiana Code section 16-39-3-10, which requires confidentiality where there is potential harm to the patient without any mention of, or concern for, the need for open court proceedings.¹⁴¹

A third factor that should be considered is the likelihood that the plaintiff will proceed with her case if anonymity is not granted as well as the likely consequences if the plaintiff abandons her claim. The plaintiff's particular community or profession may make it especially likely that the plaintiff will suffer stigmatization or damage to her professional reputation. These specific considerations may influence a plaintiff's decision whether or not to proceed with her claim, and accordingly, they should be considered by the court when

137. See *Doe v. Provident Life & Accident Ins. Co.*, 176 F.R.D. 464, 468 (E.D. Pa. 1997) (stating that the plaintiff's use of a pseudonym did not interfere with the public's right or ability to follow the proceedings and that the court intended to keep the proceedings open to the public while maintaining the plaintiff's confidentiality).

138. *Doe v. Stegall*, 653 F.2d 180, 185 (5th Cir. 1981) (internal citation omitted).

139. IND. CODE § 16-39-3-9(3).

140. See 56 C.J.S. *Mental Health* § 17.

141. IND. CODE § 16-39-3-10.

determining whether anonymity should be granted.¹⁴² Additionally, the court should consider who is responsible for compensating the plaintiff for damage caused by the defendant should the plaintiff choose to abandon her claim.¹⁴³

A fourth consideration is the possible problems the plaintiff's anonymity may create for the defendant. For example, if the plaintiff is allowed to proceed anonymously, it might be difficult for the defendant to conduct discovery and establish defenses.¹⁴⁴ Additionally, granting the plaintiff anonymity may make it impossible for the defendant to assert counterclaims and avoid the effects of res judicata.¹⁴⁵

A fifth consideration should be whether the patient's own actions regarding her mental health information should affect how a federal court treats this information. Arguably, a plaintiff who has only informed her immediate family of her illness and who has gone to great lengths to ensure that friends, colleagues, clients, and others are unaware of her disorder deserves the opportunity to continue to protect this information during litigation. By contrast, a plaintiff who has made her mental illness well-known should not necessarily get the same protection.

B. A Proposal for the Judicial Conference

Once the factors that must be considered in creating a new rule have been identified, it is necessary to determine the form this new rule should take. One plausible option is to adopt the approach used by the court in *N.Y. Blood Center*, where the plaintiff filed a complaint using her real name under seal but was allowed to use a pseudonym in the complaint in the public record.¹⁴⁶ This approach to allowing a plaintiff to proceed anonymously has many benefits because it allows the plaintiff to maintain the confidentiality of her mental health information while disclosing her identity to the party who would benefit from knowing it—the defendant. This rule would directly address the concerns cited in jurisdictions where pseudonymity is disfavored; the disclosure would prevent prejudice to the defendant in discovery as well as mitigate any risks to defendants losing their claims due to res judicata. A new rule adopting this approach would allow a plaintiff to file a complaint containing her real name under seal. With her sealed complaint, the plaintiff should be required to file a petition for leave to proceed under a pseudonym in all documents contained in the public file. The court would then treat such a petition similarly to a motion for a protective order under Federal Rule of Civil Procedure 26(c)—granting the petition when the

142. See *Provident Life & Accident Ins. Co.*, 176 F.R.D. at 468-69.

143. See, e.g., *id.* (stating that if the plaintiff chose not to pursue his claim against his insurer because of a fear of stigmatization, the federal government would have to support him financially through welfare “despite the fact that defendant may be liable to plaintiff for monthly benefit payments under the terms of the insurance policy upon which plaintiff paid premium payments”).

144. Mesler, *supra* note 50, at 882.

145. *Id.*

146. EW v. N.Y. Blood Ctr., 213 F.R.D. 108, 113 (E.D.N.Y. 2003).

party demonstrates good cause.¹⁴⁷

“Good cause,” under this approach, could be based on the five considerations discussed above. These considerations can be evaluated through the following questions, to be considered by the judge in light of the particular facts of each case:

- (1) whether the plaintiff’s mental health information will be an important part of the case, requiring that details of her mental illness be divulged;
- (2) whether an exceptional circumstance, such as danger to the public, justifies an exemption to the general rule of confidentiality of mental health information;
- (3) whether a substantial privacy interest exists such that revealing the plaintiff’s confidential mental health information may harm the plaintiff or her mental health rehabilitation process;
- (4) whether the likely harm to the plaintiff’s mental health or rehabilitation process is greater than the likely harm to the public’s right to know what transpires in the courtroom if the plaintiff is allowed to proceed anonymously;
- (5) whether the plaintiff is likely to proceed with litigation if she is not allowed to proceed anonymously, and the significance of the claim she will be abandoning if she chooses not to proceed; and
- (6) whether the plaintiff has taken care to maintain the confidentiality of her health information up to the time she filed her complaint.

When considering the answers to these questions, judges should look to HIPAA and state health information privacy statutes for help determining the weight given to the different factors and interests at stake. For example, both HIPAA and the Indiana Code balance similar interests to those at stake in determining whether a plaintiff should be allowed to proceed anonymously when her mental health information will be an important part of the litigation—privacy and public interests. HIPAA requires entities who must disclose health information to disclose the “minimum necessary for the purpose of the disclosure.”¹⁴⁸ According to HIPAA, this amount of disclosure could mean “de-identifying the information or stripping direct identifiers from the information to protect the privacy of individuals.”¹⁴⁹ Similarly, Indiana Code section 16-39-3-9 requires that disclosure of mental health records be limited “for the protection of the patient . . . and the rehabilitative process.”¹⁵⁰ These statutes privilege privacy over public interests, which is how these two interests should be balanced. Arguably, the potential effects of revealing private mental health information will generally have severe consequences for a plaintiff, while the effects of excluding the plaintiff’s name from court documents will have only minor effects on the public’s right to open access to the courts.

147. See FED. R. CIV. P. 26(c) (“The court may, for good cause, issue an order to protect a party or person from annoyance, embarrassment, oppression, or undue burden or expense.”).

148. *Health Information Privacy: Frequently Asked Questions*, *supra* note 39.

149. *Id.*

150. IND. CODE § 16-39-3-9 (2010).

Accordingly, the Judicial Conference needs to step in as soon as possible to identify and address the issues and interests at stake in litigation involving a plaintiff's mental health information. The Judicial Conference has the ability to seek input from two groups who will be directly affected by this rule—people who suffer from mental illness and the judges who will have to consider this issue in their own courtrooms. Ideally, the Judicial Conference will take the issues discussed here as a starting point from which to collect further input from the affected parties and to perfect a rule that provides protection—in the form of privacy and continued mental health—to those who need it when bringing cases in federal court.

Such a rule will help guide judges in how to weigh the private and public interests involved in determining whether a plaintiff should be allowed to proceed anonymously; thus, it leaves less room for inconsistency. Accordingly, attorneys will be able to advise their clients from the outset about how likely it is that a court will grant anonymity, which may help the client determine whether to pursue her claim in the first place (rather than after litigation has been initiated). This rule will also increase judicial efficiency in that it provides a straightforward procedure and analysis for judges to apply when mental illness is the reason a plaintiff wishes to proceed anonymously.

A new rule modeled after the one proposed above would address both the procedural and substantive inconsistencies that currently cause problems in federal courts. By requiring all plaintiffs who wish to proceed anonymously in order to protect their mental health information to file a petition requesting anonymity along with their sealed complaint, this rule simplifies the process. The rule makes it easier for litigants, attorneys, and judges to focus on the actual issue of whether the plaintiff deserves anonymity rather than whether the plaintiff complied with a particular judge's procedural preferences. Substantively, this rule puts the focus of the judge's determination on the unique implications involved when the plaintiff's privacy interest is based on her mental health information. Moreover, this rule leaves less room for a particular judge's opinions about how common a certain mental illness is or whether it is a valid cause for embarrassment or humiliation to affect whether she will grant anonymity.

CONCLUSION

The treatment of a plaintiff's ability to proceed anonymously, as it currently functions, provides protection in the form of anonymity for many groups that need it in federal civil litigation. Children, rape victims, and homosexuals are among the groups that are likely to suffer harm by revealing their names in association with the private information they reveal in the courtroom, and they are consistently granted anonymity as a result.¹⁵¹ However, there is one group that deserves to be protected which is not consistently granted the safeguard of proceeding anonymously in court. This group consists of plaintiffs who suffer

151. See *supra* notes 1-14 and accompanying text.

from mental illness—or who have suffered from a mental illness in the past—and wish to bring a cause of action which will require that the details of their mental health be disclosed and examined by the court and the opposing party, possibly in great detail.

Most federal courts addressing whether a plaintiff may proceed anonymously apply some form of multi-factor test to weigh the litigant's privacy interest against the public interest in open court proceedings. However, judges in different federal circuits weigh the various factors unevenly. Thus, while most courts generally apply a multi-factor test, there remains a significant lack of consistency among circuits in how plaintiffs whose claims involve their mental health information will be treated in the courtroom.¹⁵²

This lack of consistency is problematic because mental illness is still highly stigmatized in our society.¹⁵³ Being labeled “mentally ill” often makes it more difficult to obtain employment and housing, and it can have negative repercussions on professional and social relationships.¹⁵⁴ Consequently, plaintiffs who suffer or have suffered from a mental health disorder often have legitimate reasons for wishing to keep their mental health information private. This significant privacy interest and the lack of consistency in whether plaintiffs are granted anonymity when their mental health information will be revealed in the course of litigation often means that plaintiffs are left with two unacceptable options. First, the plaintiff can proceed with litigation and divulge the details of her mental health information, risking stigmatization and social or professional consequences. Second, the plaintiff can abandon her claim in order to protect very personal mental health information, thereby forgoing the opportunity to pursue what may well be a valid claim. In determining whether to grant a particular plaintiff's motion to proceed anonymously, a court should have the flexibility to consider the potential negative repercussions of denying the motion as well as the plaintiff's reasons for wanting to keep this information private.

The common law rule that has evolved in the federal courts to determine whether a plaintiff should be allowed to proceed anonymously when mental health information is involved has produced jumbled results that are inconsistent with our country's usual approach to health information. Federal and state health information privacy statutes provide safeguards to protect citizens' medical records and demonstrate the very high value we generally place on privacy interests related to health information.¹⁵⁵ However, when it comes to allowing a plaintiff to proceed anonymously in a case that involves her mental health information, courts in some jurisdictions are quick to find that the public interest in open court proceedings outweighs the litigant's privacy interest.

But why is personal health information that comes out in the course of litigation any less worthy of protection than the health information contained in a person's medical records and explicitly protected by federal and state health

152. See discussion *supra* Part II.B.

153. Teachman, *supra* note 14, at 77.

154. See discussion *supra* Part I.A.

155. See discussion *supra* Part I.B.

information privacy statutes? Because the same privacy interests are involved, it only makes sense to treat medical information that comes out in litigation the same way we treat medical information in other circumstances.

The best policymaking body to take on the task of creating a new federal rule making the treatment of health information confidentiality consistent across forums is the Judicial Conference of the United States.¹⁵⁶ Accordingly, the Judicial Conference should adopt a new set of procedures, rules, and factors for courts to weigh in determining whether a plaintiff should be allowed to proceed anonymously when her mental health information is involved. Such procedures, rules, and factors should consider the specific implications of mental illness and the sensitivity of a plaintiff's mental health information as well as the realities of judicial "openness."

A central tenet of this new rule should be that the right of plaintiffs to maintain confidentiality regarding their mental illnesses by proceeding anonymously and the right of the public to open trials are not necessarily as dichotomous as they are made out to be. Thus, the rights and interests of both plaintiffs and the public can and should be protected by a new rule that recognizes the unique implications in issues as private and sensitive as mental illness. The rule proposed in this Note does precisely that by allowing plaintiffs to file their complaints under seal and to proceed using a pseudonym in the public file. This approach protects the plaintiff's privacy and the confidentiality of her mental health information—consistent with federal and state health information privacy statutes—and protects the public's right to know what transpires in the courtroom. Ultimately, the system is preserved because all court proceedings remain public; only the plaintiff's name is withheld.

156. See discussion *supra* Part II.A.

“GROWING PAINS” IN INDIANA AGE DISCRIMINATION LAW

DAVID T. VLINK*

INTRODUCTION

Age discrimination is on the rise. Age discrimination claims are now the fastest-growing category of employment discrimination claims being filed with the Equal Employment Opportunity Commission (EEOC).¹ From 2007 to 2008, age discrimination claims increased nearly 29%, more than doubling the increase in claims for sex discrimination (14%), and nearly tripling the increase in claims for race discrimination (11%).² In fact, in terms of the total number of charges filed with the EEOC, age discrimination is the third most frequently charged form of employment discrimination behind race and sex discrimination.³ As these numbers indicate, age discrimination is a very real and increasing problem. As a result, Indiana must take steps to strengthen its protections against age discrimination in the workplace.

The Supreme Court recently issued a landmark age discrimination ruling in *Gross v. FBL Financial Services, Inc.*,⁴ which brought the issue of age discrimination to the forefront of employment discrimination law. In *Gross*, the Supreme Court held that even when an age discrimination plaintiff shows that age was a “motivating factor” in an employer’s employment decision, the burden never shifts to the employer to prove that it would have made the same decision regardless of the plaintiff’s age.⁵ Consequently, a mixed-motive jury instruction is never proper in employment discrimination cases brought pursuant to the federal Age Discrimination in Employment Act (ADEA).⁶ The *Gross* decision means that from now on, absent Congressional action,⁷ age discrimination

* J.D. Candidate, Indiana University School of Law—Indianapolis, 2011; B.A., Purdue University Calumet, Hammond, Indiana, 2007.

1. See U.S. EQUAL OPPORTUNITY COMM’N, CHARGE STATISTICS FY 1997 THROUGH FY 2010, <http://www.eeoc.gov/eeoc/statistics/enforcement/charges.cfm> (last visited Feb. 3, 2011) [hereinafter EEOC, CHARGE STATISTICS].

2. According to the EEOC, the following changes occurred from 2007 to 2008: (1) age discrimination charges increased from 19,103 to 24,582; (2) sex-related charges increased from 24,826 to 28,372; and (3) race-related charges increased from 30,510 to 33,937. EEOC, CHARGE STATISTICS, *supra* note 1.

3. *See id.*

4. 129 S. Ct. 2343 (2009).

5. *Id.* at 2352.

6. 29 U.S.C. §§ 621-34 (2006 & Supp. 2009).

7. Immediately after the *Gross* decision came down, Congressman George Miller (D-Cal.), Chairman of the House Committee on Education and Labor, issued a press release stating that his committee would hold hearings on the *Gross* decision, with the purpose of overturning it. *See* Press Release, H. Comm. on Educ. and Labor, Congress to Hold Hearings on Supreme Court’s “Gross” Ruling on Age Discrimination, Says Chairman Miller (June 30, 2009), *available at*

plaintiffs must prove that age was the “but-for” cause of the alleged discriminatory employment action.⁸ Commentators have noted that the “but-for” standard is “a stricter and typically more difficult showing for the plaintiff” and will make it more difficult for age discrimination plaintiffs to prevail on their claims.⁹

This recent federal development in age discrimination law is particularly relevant to Indiana employees because the Indiana Age Discrimination Act (IADA)¹⁰ specifically exempts from its coverage any employer subject to the ADEA.¹¹ The ADEA applies to all employers nationwide with twenty or more employees.¹² Thus, most Indiana employees who allege age discrimination must proceed under the ADEA, with its recently restricted burden of proof, because there is no comparable state statute under which the employee can seek relief. Moreover, even for workers who are covered because their employer has fewer than twenty employees, the IADA does not provide them a private right of action to take their claims to court. Instead, the IADA leaves enforcement entirely to the Indiana Commissioner of Labor, who has little remedial authority.¹³ In addition, a recent Indiana Supreme Court decision, *Montgomery v. Board of Trustees of Purdue University*,¹⁴ held that state employers are “subject to” the ADEA.¹⁵ This holding came despite the United States Supreme Court’s ruling in *Kimel v.*

<http://democrats.edworkforce.house.gov/newsroom/2009/06/congress-to-hold-hearing-on-su.shtml> (last visited Feb. 3, 2011). Subsequently, on September 6, 2009, Congressman Miller—along with Senators Tom Harkin (D-Iowa) and Pat Leahy (D-Vt.)—introduced the Protecting Older Workers Against Age Discrimination Act, H.R. 3721, 111th Cong. (2009), which would overturn *Gross* and make clear that the mixed-motive motivating factor standard applies to ADEA cases. As of this writing on February 3, 2011, however, no major action has been taken on this legislation.

8. *Gross*, 129 S. Ct. at 2352.

9. E.g., Jeffrey Campolongo, *Gross Decision May Result in More Older Workers Being Fired*, *LEGAL INTELLIGENCER*, June 26, 2009 (contending that the *Gross* decision “certainly raises the burden for ADEA plaintiffs”).

10. IND. CODE §§ 22-9-2-1 to -11 (2010).

11. See *id.* § 22-9-2-1 (noting that the term “employer” does not include “a person or governmental entity which is subject to the federal Age Discrimination in Employment Act”).

12. 29 U.S.C. § 630(b) (2006) (“The term ‘employer’ means a person engaged in an industry affecting commerce who has twenty or more employees for each working day in each of twenty or more calendar weeks in the current or preceding calendar year.”).

13. See *Hague v. Thompson Distrib. Co.*, No. 1:02-CV-01744-RLY-TA, 2005 WL 995689, at *7 (S.D. Ind. Feb. 9, 2005). The *Hague* court explained that

[t]here is no private right of action provided under state law. A complaining party must take up its age discrimination complaints with the Commissioner of Labor, who has the power to investigate, facilitate, and hold hearings on the complaint. Other than that, the . . . [IADA] offers no remedy to an individual.

Id. (citing *Helman v. AMF, Inc.*, 675 F. Supp. 1163, 1165 (S.D. Ind. 1987)).

14. 849 N.E.2d 1120 (Ind. 2006).

15. See *id.* at 1126-28.

*Florida Board of Regents*¹⁶ that the Eleventh Amendment¹⁷ bars age discrimination actions for damages by state employees against their employers in federal court.¹⁸ As a result of *Montgomery*, state employees have no protection under the IADA; as a result of *Kimel*, they also cannot pursue their claims in federal court. The effect of these statutes and other precedent leaves Indiana employees inadequately protected from age discrimination—particularly state government employees, who find themselves in a “Catch-22” situation where they are unable to seek redress in either the state or federal courts.

The purpose of this Note is twofold: to demonstrate that Indiana does not adequately protect its workers from age discrimination and to propose solutions to this problem. Part I examines Indiana’s treatment of age discrimination, focusing mainly on the IADA and the *Montgomery* decision, and contrasting it with how Indiana treats other forms of discrimination in the Indiana Civil Rights Law (ICRL).¹⁹ Part II addresses federal treatment of age discrimination before *Gross*, given its particular relevance to Indiana workers. Part III looks at how other states treat age discrimination, with an emphasis on Indiana’s neighboring states. Part IV proposes solutions to the problem, drawing on the approaches of other states and previous attempts at reform proposed by Indiana legislators.

I. INDIANA’S TREATMENT OF AGE DISCRIMINATION

This section examines Indiana’s treatment of age discrimination by introducing and analyzing the Indiana Age Discrimination Act (IADA) and *Montgomery*, a recent Indiana Supreme Court case that has had a major impact on how the IADA is now interpreted in Indiana. Additionally, this section takes a brief look at administrative treatment of age discrimination under the Indiana State Personnel Act.²⁰ Finally, this section concludes by contrasting the IADA with Indiana’s primary civil rights statute, the ICRL.

A. The Indiana Age Discrimination Act

Under Indiana law, age discrimination in employment is considered an “unfair employment practice and against public policy.”²¹ Specifically, the IADA provides:

16. 528 U.S. 62 (2000).

17. U.S. CONST. amend. XI. The Eleventh Amendment, which states that “[t]he Judicial power of the United States shall not be construed to extend to any suit in law or equity, commenced or prosecuted against one of the United States by Citizens of another State, or by Citizens or Subjects of any Foreign State[,]” has generally been interpreted to prohibit suits against states for damages in federal courts. *See Hans v. Louisiana*, 134 U.S. 1, 10-11 (1890).

18. *Kimel*, 528 U.S. at 91 (holding that “in the ADEA, Congress did not validly abrogate the States’ sovereign immunity to suits by private individuals.”).

19. IND. CODE §§ 22-9-1-1 to -18 (2010).

20. *Id.* §§ 4-15-2-1 to -43.

21. Karl Oakes, 5 IND. LAW ENCYC. *Civil Rights* § 11 (West 2011).

It is declared to be an unfair employment practice and to be against public policy to dismiss from employment, or to refuse to employ or rehire, any person solely because of his age if such person has attained the age of forty (40) years and has not attained the age of seventy-five (75) years.²²

The IADA, however, does not provide victims of age discrimination in employment a private cause of action to vindicate their rights in court.²³ Instead, remedial authority is left solely to the Indiana Commissioner of Labor ("the Commissioner"), who the IADA charges with investigating complaints of age discrimination and combating such discrimination through "informal methods of conference, conciliation and persuasion."²⁴ Despite the Commissioner's apparent authority, the Commissioner is effectively powerless to combat age discrimination. The most the Commissioner can do is "state his findings of fact" that an employer committed age discrimination.²⁵ Thus, although the stated policy of the IADA is clear and forceful, its remedial effect is minimal.

In addition to the IADA's lack of remedies, the statute also has a coverage problem. Section 1 of the IADA states, "'Employer' shall mean and include any person in this state employing one (1) or more individuals . . . the state and all political subdivisions, boards . . . and commissions thereof, but does not include . . . a person or governmental entity which is subject to the federal Age Discrimination in Employment Act."²⁶ The effect of this provision is that the Commissioner and Indiana state courts are "deprived of jurisdiction if the employer falls within the definition of employer as defined under [the] ADEA."²⁷ Given that the ADEA applies to any employer nationwide with twenty or more employees,²⁸ the IADA only applies to a small number of employers: those with fewer than twenty employees.

B. Montgomery v. Board of Trustees of Purdue University

There have been only a small number of Indiana cases dealing with age discrimination. Presumably, most Indiana age discrimination plaintiffs choose to pursue their claims under the federal statute. However, one recent Indiana Supreme Court case—*Montgomery v. Board of Trustees of Purdue University*²⁹—addresses age discrimination under the IADA.

22. IND. CODE § 22-9-2-2.

23. *Hague v. Thompson Distrib. Co.*, No. 1:02-CV-01744-RLY-TA, 2005 WL 995689, at *7 (S.D. Ind. Feb. 9, 2005) (explaining that "[t]here is no private right of action provided . . . [in the IADA].").

24. IND CODE §§ 22-9-2-5 to -6.

25. *Id.* § 22-9-2-6.

26. *Id.* § 22-9-2-1.

27. *Helman v. AMF, Inc.*, 675 F. Supp. 1163, 1165 (S.D. Ind. 1987).

28. 29 U.S.C. § 630(b) (2006).

29. 849 N.E.2d 1120 (Ind. 2006).

In *Montgomery*, plaintiff Michael Montgomery had worked for Purdue University (“the University”) for nearly thirty years before the University fired him in 2002 at the age of fifty-eight.³⁰ The University terminated Montgomery for declining involuntary retirement³¹ and conceded, for the purposes of a motion to dismiss, that it terminated Montgomery solely because of his age.³² Montgomery filed a complaint in state court, bypassing the Commissioner. Both lower courts granted the University’s motion to dismiss for failure to state a claim.³³

On appeal to the Indiana Supreme Court, the issue was whether the University was “subject to” the ADEA and therefore exempt from the IADA.³⁴ The Indiana Supreme Court began its analysis by noting that the ADEA has two primary enforcement mechanisms: direct enforcement by the Equal Employment Opportunity Commission (EEOC) and private civil actions by aggrieved individuals.³⁵ The court then discussed the doctrine of sovereign immunity under the Eleventh Amendment.³⁶ It particularly focused on the United States Supreme Court’s decision in *Kimel v. Florida Board of Regents*,³⁷ which held that the ADEA’s provisions authorizing state employees to sue their employers in federal court for money damages violated the Eleventh Amendment.³⁸ The court noted that pursuant to *Kimel*, the University was clearly not “subject to” the ADEA under the private litigant method of enforcement in federal courts, as Indiana has not consented to ADEA suits in federal court.³⁹ The court pointed out, however, that pursuant to the doctrine of *Ex parte Young*,⁴⁰ an aggrieved state employee can sue a state employer for injunctive relief in federal court by naming a state official as the defendant. Furthermore, the EEOC, as an arm of the federal government, is not subject to the Eleventh Amendment’s prohibition on suing states in federal court.⁴¹

Montgomery argued that these remedies were “meaningless” and “illusory” because the EEOC rarely exercises its direct enforcement authority, and the decision to grant the injunctive remedy of reinstatement—albeit subject to the

30. See *id.* at 1122.

31. *Montgomery v. Bd. of Trs. of Purdue Univ.*, 824 N.E.2d 1278, 1279 (Ind. Ct. App. 2005), aff’d in part, 849 N.E.2d 1120 (Ind. 2006).

32. Brief of Appellant at 2, *Montgomery v. Bd. of Trs. of Purdue Univ.*, 849 N.E.2d 1278 (Ind. Ct. App. 2005) (No. 79A05-0411-CV-591), 2004 WL 3216683 at *2.

33. *Montgomery*, 824 N.E.2d at 1279, 1282-83.

34. *Montgomery*, 849 N.E.2d at 1122-23.

35. *Id.* at 1123.

36. U.S. CONST. amend XI.

37. 528 U.S. 62 (2000).

38. *Id.* at 91.

39. *Montgomery*, 849 N.E.2d at 1125-26. The court noted that Indiana has not consented to ADEA suits in federal courts because waiver of sovereign immunity by the states must be “express, unequivocal and voluntary.” *Id.* (citing *Edelman v. Jordan*, 415 U.S. 651, 673 (1974)).

40. 209 U.S. 123 (1908).

41. *Montgomery*, 849 N.E.2d at 1126-27.

availability of positions—is within the trial court judge’s discretion.⁴² The court ultimately rejected Montgomery’s arguments, concluding that “[i]f the law imposes standards of conduct on state employers, they are ‘subject to’ it. The fact that some remedies may be constitutionally barred does not change this result.”⁴³ Thus, the court held that Indiana state employers are “subject to” the ADEA; as such, they are not statutory employers under section 1 of the IADA.⁴⁴ Post-*Montgomery*, therefore, state government employees cannot bring IADA claims because state employers are “subject to” the ADEA.

After finding the University “subject to” the ADEA, the court next addressed Montgomery’s argument that the IADA creates a private cause of action for violation of public policy under *Frampton v. Central Indiana Gas Co.*⁴⁵ In *Frampton*, the Indiana Supreme Court created an exception to the employment-at-will doctrine⁴⁶ by holding that an employee has a claim for wrongful discharge “when . . . [the] employee is discharged solely for exercising a statutorily conferred right.”⁴⁷ The court rejected this argument for multiple reasons. First, the court explained that “the IADA has never expressly provided for enforcement through private judicial action”⁴⁸ and that “the legislative history of the IADA does not support Montgomery’s argument that the [g]eneral [a]ssembly intended to create a private cause of action for monetary damages under the IADA.”⁴⁹ The court further supported its finding by opposing the use of a public policy argument, stating that “[g]eneral expressions of public policy do not support new exceptions to the employment-at-will doctrine.”⁵⁰ Thus, the court ultimately concluded that the text of the IADA, coupled with its legislative history and the court’s own jurisprudence on employment-at-will, weighed against creating a private cause of action for age discrimination in Indiana. In sum, whether Indiana employees who have been discriminated against because of their age should have any remedies “is a policy judgment” which “must come from the [g]eneral [a]ssembly.”⁵¹

Justice Rucker authored a short dissent in *Montgomery* in which he characterized Montgomery’s situation as a “Catch-22,” stating, “So here we have an employee fired because of his age but [who] in effect has no remedy, according to today’s opinion, despite both state and federal legislation designed to protect employees fired because of age. Certainly the legislature could not

42. *Id.* at 1127.

43. *Id.*

44. *Id.*

45. 297 N.E.2d 425 (Ind. 1973).

46. In Indiana, the employment-at-will doctrine states that either an employee or employer may terminate the employment relationship at any time for “good reason, bad reason, or no reason at all.” *Cantrell v. Morris*, 849 N.E.2d 488, 494 (Ind. 2006).

47. *Frampton*, 297 N.E.2d at 428.

48. *Montgomery*, 849 N.E.2d at 1129.

49. *Id.* at 1128.

50. *Id.*

51. *Id.* at 1130-31.

have intended the result reached in this case.”⁵² Unfortunately, Justice Rucker did not provide any analysis to accompany his opinion. There is, however, some circumstantial evidence to bolster Justice Rucker’s claim. In *Town of South Whitley v. Cincinnati Insurance Co.*,⁵³ the court stated that the IADA “is intended to provide coverage only where a plaintiff cannot proceed under the federal act.”⁵⁴ This makes sense considering that the provision of the IADA exempting employers “subject to” the ADEA was enacted pre-*Kimel*, when it was thought that Congress had broad authority to abrogate state sovereign immunity under Section 5 of the Fourteenth Amendment.⁵⁵ Furthermore, even if one accepts the court’s reasoning in *Montgomery* that state employers are still “subject to” the ADEA, there is no doubt that individual plaintiffs are severely restricted in their ability to proceed under the ADEA post-*Kimel*. In any event, Justice Rucker’s ultimate conclusion was correct: Montgomery was left without a remedy, as the court dismissed his lawsuit against the University for failure to state a claim.⁵⁶

Montgomery was not the only person adversely affected by this ruling, though. In *Keene v. Marion County Superior Court*,⁵⁷ the plaintiff, a sixty-four-year-old man, was terminated by his employer and replaced with a forty-year-old employee.⁵⁸ He filed a complaint for age discrimination under the IADA, which the lower courts initially dismissed as untimely.⁵⁹ On appeal to the Indiana Supreme Court, the issue was whether the lower courts were correct in deeming the complaint time-barred.⁶⁰ The court did not reach this issue, however, noting that “[s]ubsequent to our granting transfer and holding oral arguments in this case, we decided *Montgomery*”⁶¹ Thus, the court held, “Because the employer here is not subject to the IADA (because it is subject to the federal ADEA), the trial court correctly dismissed Keene’s claims.”⁶² As in *Montgomery*, a government employee allegedly fired for age discrimination “in effect . . . [had] no remedy.”⁶³

C. The Indiana State Personnel Act

Lest one get the impression that Indiana state government employers are free to terminate their older workers with impunity, it bears noting that the Indiana

52. *Id.* at 1131 (Rucker, J., dissenting).

53. 724 F. Supp. 599 (N.D. Ind. 1989).

54. *Id.* at 603 (citing *Helman v. AMF, Inc.*, 675 F. Supp. 1163 (S.D. Ind. 1987)).

55. U.S. CONST. amend. XIV, § 5.

56. *Montgomery*, 849 N.E.2d at 1131 (majority opinion).

57. 849 N.E.2d 1141 (Ind. 2006).

58. *Id.* at 1141.

59. *Id.* at 1141-42.

60. *Id.* at 1142.

61. *Id.*

62. *Id.*

63. *Montgomery v. Bd. of Trs. of Purdue Univ.*, 849 N.E.2d 1120, 1131 (Ind. 2006) (Rucker, J., dissenting).

State Personnel Act (“Personnel Act”)⁶⁴ prevents them from doing so. The Personnel Act prohibits state employers from discriminating against their employees “on the basis of politics, religion, sex, age, race, or because of membership in an employee organization.”⁶⁵ The Personnel Act also sets up a system whereby state employees who allege mistreatment can file a complaint with the head of the agency for which they work and, if the head’s decision is unsatisfactory, with the Indiana State Personnel Department’s personnel board.⁶⁶ If the employee is unsatisfied with the personnel board’s findings, the employee can appeal to the Indiana State Employee Appeals Commission (SEAC).⁶⁷ If the SEAC finds discrimination, “the employee shall be reinstated without loss of pay.”⁶⁸ But the SEAC’s authority to fashion remedies is limited to reinstatement, and it cannot order an agency to create a new position if the employee’s position is no longer available.⁶⁹

The Personnel Act does not, however, cover every employee who would be considered a state employee for Eleventh Amendment purposes. Purdue University, for example, is not listed as part of the “state service” in section 3.8 of the Personnel Act.⁷⁰ Moreover, Montgomery would not appear to be a “regular employee” as defined in section 3.7 of the Personnel Act (as is required in order to file a complaint with the Indiana State Personnel Department).⁷¹ Therefore, even though the Personnel Act could be used as a fallback remedy for certain employees wishing to pursue reinstatement (and only reinstatement) for alleged age discrimination, it appears that Montgomery also would not have been able to proceed under the Personnel Act.

D. The IADA Compared to the Indiana Civil Rights Law

The Indiana Civil Rights Law (ICRL)⁷² is Indiana’s primary anti-

64. IND. CODE §§ 4-15-2-1 to -43 (2010).

65. *Id.* § 4-15-2-35 (emphasis added).

66. *Id.* The complaint must be filed within thirty (30) days of the alleged mistreatment, and it must be in writing. *Id.* The Indiana State Personnel Department is a state agency charged with administering the Personnel Act. *Id.* § 4-15-1.8-7(a)(18).

67. *Id.* § 4-15-2-35. The SEAC is a bipartisan board whose mission is “[t]o hear or investigate those appeals from state employees as is set forth in . . . [the Personnel Act], and fairly and impartially render decisions as to the validity of the appeals or lack thereof.” *Id.* § 4-15-1.5-6(1).

68. *Id.* § 4-15-2-35.

69. See *Ind. Dep’t of Env’tl. Mgmt. v. West*, 838 N.E.2d 408, 417 (Ind. 2005). In *West*, the court applied the federal disparate treatment framework of *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973), during the liability phase in reviewing the SEAC’s finding that the Indiana Department of Environmental Management (IDEM) violated the Personnel Act. *West*, 838 N.E.2d at 413.

70. See IND. CODE § 4-15-2-3.8.

71. See *id.* §§ 4-15-2-3.7 to -3.8.

72. *Id.* §§ 22-9-1-1 to -18.

discrimination statute. The law prohibits discrimination based on “race, religion, color, sex, disability, national origin, or ancestry,”⁷³ but not age. The ICRL states that “equal . . . employment opportunities . . . are hereby declared to be civil rights” and that “[i]t is the public policy of the state to provide all of its citizens equal opportunity for . . . employment.”⁷⁴ Furthermore, the ICRL provides that

[t]he practice of denying these rights [equal education, employment, access to public accommodations, and opportunity for acquisition of land] to properly qualified persons by reason of the race, religion, color, sex, disability, national origin, or ancestry of such person is contrary to the principles of freedom and equality of opportunity and is a burden to the objectives of public policy of this state and shall be considered as discriminatory practices.⁷⁵

The ICRL is significantly more comprehensive in scope and remedies than the IADA. There are several major discrepancies in the way Indiana treats age discrimination under the IADA compared to other forms of discrimination under the ICRL, the most important of which are discussed below.

1. *Declaration of the Right.*—The first meaningful distinction between the IADA and the ICRL is that the ICRL declares equal opportunities for each of the listed classes—race, religion, color, sex, disability, national origin, and ancestry—to be a “civil right,”⁷⁶ whereas the IADA merely declares age discrimination to be “unfair.”⁷⁷

2. *Investigating Complaints.*—The ICRL sets up an independent Indiana Civil Rights Commission, which is authorized to “receive and investigate complaints alleging discriminatory practices.”⁷⁸ The Indiana Civil Rights Commission is further empowered to hold hearings on complaints of discrimination and to fashion meaningful remedies in the event that it finds discrimination to have occurred in a given case.⁷⁹ The IADA, on the other hand, puts the Indiana Commissioner of Labor in charge of investigating complaints as an ancillary responsibility.⁸⁰ This discrepancy in treatment sheds light on the level of importance Indiana puts on eliminating age discrimination as compared to the other types of discrimination covered by the ICRL.

3. *Remedies.*—The remedies available under the ICRL are far superior to the remedies—or lack thereof—available under the IADA. Under the ICRL, the Indiana Civil Rights Commission is authorized to issue cease and desist orders

73. *Id.* § 22-9-1-3(l)(1)-(2).

74. *Id.* § 22-9-1-2(a).

75. *Id.* § 22-9-1-2(b). Thus, similar to federal law, but unlike any of its neighboring states, Indiana fails to make age a protected class under its primary anti-discrimination statute. However, unlike federal law, Indiana includes disability as a protected class.

76. *Id.* § 22-9-1-2(a).

77. Compare IND. CODE § 22-9-1-2, with *id.* § 22-9-2-2.

78. *Id.* § 22-9-1-6(e).

79. See *id.* § 22-9-1-6(k)(A)-(D).

80. *Id.* §§ 22-9-2-5 to -6.

and orders requiring violators to “restore [the] complainant’s losses incurred as a result of the discriminatory treatment, as the commission may deem necessary to assure justice.”⁸¹ Restoration of the victim’s losses includes the equitable remedy of reinstatement⁸² and economic damages to compensate for lost wages, salary, or commissions.⁸³ In stark contrast, the IADA leaves enforcement entirely up to the Indiana Commissioner of Labor as an ancillary responsibility, and the most he can do is “state his findings of fact.”⁸⁴

4. *Availability of a Private Cause of Action.*—The *Montgomery* court determined that the IADA does not allow an age discrimination plaintiff to file a lawsuit against the alleged discriminator.⁸⁵ The ICRL, on the other hand, specifically provides for a private right of action.⁸⁶ Section 16 of the ICRL states that “[a] respondent or a complainant may elect to have the claims that are the basis for a finding of probable cause decided in a civil action However, both the respondent and complainant must agree in writing to have the claims decided in a court of law.”⁸⁷ This option is not available, however, if the Indiana Civil Rights Commission has already begun a hearing on the alleged discrimination.⁸⁸ If the case is decided in court and the plaintiff proves discrimination, the court can grant the same remedies as the Indiana Civil Rights Commission.⁸⁹ Even though the ability to pursue a private right of action under the ICRL is limited by provisions requiring both parties to agree and eliminating the right if the commission has begun a hearing,⁹⁰ the availability of a private right of action under the ICRL is important. The private right of action is recognized as an effective means of redressing grievances.⁹¹ In sum, the fact that the ICRL provides for a private right of action means that it provides more protection than the IADA, despite the fact that the ICRL imposes some limitations on a plaintiff’s

81. *Id.* § 22-9-1-6(k)(A).

82. See *Ind. Civil Rights Comm’n v. Culver Educ. Found.*, 510 N.E.2d 206, 211 (Ind. Ct. App. 1987) (approving the Commission’s order granting reinstatement since “the remedy of reinstatement is not precluded by [section 6(k)(A) of the ICRL]”), vacated by 535 N.E.2d 112 (Ind. 1989); *Ind. Civil Rights Comm’n v. Midwest Steel Div.*, 450 N.E.2d 130, 140 (Ind. Ct. App. 1983) (noting that “the remedial power of reinstatement . . . [was] not precluded.”).

83. See *Midwest Steel Div.*, 450 N.E.2d at 140 (noting that economic damages for employment discrimination are permitted, but emotional damages are precluded).

84. IND CODE §§ 22-9-2-5 to -6.

85. *Montgomery v. Bd. of Trs. of Purdue Univ.*, 849 N.E.2d 1120, 1130-31 (Ind. 2006).

86. See IND. CODE § 22-9-1-16(a).

87. *Id.*

88. *Id.* § 22-9-1-16(b).

89. *Id.* § 22-9-1-17(b) (“If the court finds that a discriminatory practice has occurred[,] the court may grant the relief allowed under . . . [Indiana Code section] 22-9-1-6(k).”).

90. See Kathryn E. Olivier, Note, *The Effect of Indiana Code Section 22-9-1-16 on Employee Civil Rights*, 42 IND. L. REV. 441, 467 (2009) (arguing that Indiana should not require consent from both parties to institute a private civil action).

91. See 42 U.S.C. § 1983 (2006) (providing a private cause of action for constitutional violations).

ability to file a discrimination lawsuit.

II. FEDERAL TREATMENT OF AGE DISCRIMINATION

In *Gross v. FBL Financial Services, Inc.*,⁹² the United States Supreme Court recently altered the standard for proving disparate treatment claims under the ADEA. Many commentators believe this changed standard makes it more difficult for ADEA plaintiffs to succeed on their claims.⁹³ Because of the IADA’s provision exempting employers “subject to” the ADEA, this recent federal development is particularly relevant to Indiana employees’ discrimination claims. This section will provide a brief history of federal anti-discrimination law up to *Gross* and give a detailed analysis of the *Gross* decision and its implications for future ADEA claims.

A. Federal Anti-Discrimination Law Before Gross

Title VII of the Civil Rights Act of 1964 (“Title VII”)⁹⁴ is the grandparent of employment discrimination statutes. Although the ultimate goal of Title VII is to eliminate discrimination in the workplace, Title VII was primarily proposed to combat racial discrimination in employment against blacks.⁹⁵ As enacted, however, Title VII prohibited discrimination on the basis of color, sex, national origin, or religion. Section 703(a)(2), the primary anti-discrimination rule of Title VII, states:

It shall be an unlawful employment practice for an employer . . . to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment because of such individual’s race, color, religion, sex, or national origin⁹⁶

In the debates leading up to the passage of Title VII, Congress considered adding age as a protected category, but that idea was ultimately rejected.⁹⁷ Congress felt that it did not know enough about age discrimination and ordered the Secretary of Labor (“the Secretary”) “to make a full and complete study of the factors which might tend to result in discrimination in employment because of age and of the consequences of such discrimination on the economy and

92. 129 S. Ct. 2343 (2009).

93. See, e.g., Campolongo, *supra* note 9 (stating that the new standard announced in *Gross* is “a stricter and typically more difficult showing for the plaintiff”; thus, “the *Gross* decision “certainly raises the burden for ADEA plaintiffs”).

94. 42 U.S.C. §§ 2000e-2000e-17 (2006 & Supp. 2009).

95. See *United Steelworkers of Am. v. Weber*, 443 U.S. 193, 202 (1979) (commenting that “the goal[] of the Civil Rights Act . . . [was] the integration of blacks into the mainstream of American society . . . ”).

96. 42 U.S.C. § 2000e-2(a)(1).

97. See *Smith v. City of Jackson*, 544 U.S. 228, 232 (2005) (noting that amendments offered to add age as a protected class under Title VII were twice defeated).

individuals affected.”⁹⁸ The Secretary issued his report in June 1965 and found that although age discrimination was generally not prejudice-based, like race discrimination, there was a significant problem of discrimination against older workers which involved “their rejection because of assumptions about the effect of age on their ability to do the job *when there is in fact no basis for these assumptions.*”⁹⁹ The Secretary urged Congress to enact “effective measures in this most deserving and much neglected cause.”¹⁰⁰

Congress followed the Secretary’s advice and passed the ADEA in 1967.¹⁰¹ In passing the ADEA, Congress sought to “promote employment of older persons based on their ability rather than age” and to “prohibit arbitrary age discrimination in employment.”¹⁰² As enacted, the ADEA used identical “because of” language to prohibit age discrimination as was used in Title VII to prohibit other forms of discrimination.¹⁰³ Borrowing from the language used in Title VII, the primary anti-discrimination rule of the ADEA states, “It shall be unlawful for an employer . . . to limit, segregate, or classify his employees in any way which would deprive . . . any individual of employment opportunities or otherwise adversely affect his status as an employee, because of such individual’s age . . . ”¹⁰⁴

Other than making it illegal to discriminate against an individual “because of” a protected class, Congress provided little guidance to courts on how to analyze disparate treatment claims under both Title VII and the ADEA. The courts thus have evaluated disparate treatment claims under the burden-shifting framework of *McDonnell Douglas Corp. v. Green*.¹⁰⁵ Proving discrimination under *McDonnell Douglas* and its progeny involves a three step burden-shifting

98. Civil Rights Act of 1964, Pub. L. No. 88-352, § 715, 78 Stat. 241, 265.

99. REPORT OF THE SEC’Y OF LABOR, 89TH CONG., THE OLDER AMERICAN WORKER: AGE DISCRIMINATION EMPLOYMENT 2 [hereinafter REPORT OF THE SEC’Y OF LABOR] (emphasis in original).

100. *Id.* at 1.

101. Age Discrimination in Employment Act of 1967, Pub. L. No. 90-202, 81 Stat. 602 (codified at 29 U.S.C. §§ 621-34).

102. 29 U.S.C. § 621(b).

103. Compare 29 U.S.C. § 623(a) (declaring it unlawful to discriminate “because of” age), with 42 U.S.C. § 2000e-2(a)(1) (declaring it unlawful to discriminate “because of” race, color, religion, sex, or national origin).

104. 29 U.S.C. § 623(a)(2).

105. 411 U.S. 792 (1973). Courts have traditionally applied the *McDonnell Douglas* framework to ADEA claims. See, e.g., *Reeves v. Sanderson Plumbing Prods., Inc.*, 530 U.S. 133, 141 (2000) (noting that the courts of appeals have employed *McDonnell Douglas* in ADEA disparate treatment claims and assuming that it applied there). In *Gross*, however, the Supreme Court cast doubt on *McDonnell Douglas*’s continuing applicability to ADEA claims, stating that “the Court has not definitively decided whether the evidentiary framework of *McDonnell Douglas* . . . is appropriate in the ADEA context.” *Gross v. FBL Fin. Servs., Inc.*, 129 S. Ct. 2343, 2349 n.2 (2009).

process. First, a plaintiff must make a *prima facie* showing of discrimination.¹⁰⁶ Second, if the plaintiff successfully makes a *prima facie* showing, a presumption of discrimination arises, and the employer has the burden of articulating a legitimate non-discriminatory reason for its actions.¹⁰⁷ Third, once the employer articulates its legitimate reason, the plaintiff has an opportunity to rebut the employer’s reason by showing that it is pretextual.¹⁰⁸ If the plaintiff proves pretext, then the jury is allowed to infer discrimination.¹⁰⁹ This process is known as the “pretext method”¹¹⁰ of proving disparate treatment.

Another way to prove disparate treatment is to utilize the “mixed-motives method.”¹¹¹ Initially endorsed in *Price Waterhouse v. Hopkins*,¹¹² the mixed-motives method allows a plaintiff to prevail under disparate treatment by showing that a protected trait was a “motivating factor” in the challenged employment action.¹¹³ As in the pretext situation, a mixed-motives plaintiff is still required to

106. *McDonnell Douglas*, 411 U.S. at 802. To make a *prima facie* case, a plaintiff must show (i) that he belongs to a . . . [protected class]; (ii) that he applied and was qualified for a job for which the employer was seeking applicants; (iii) that, despite his qualifications, he was rejected; and (iv) that, after his rejection, the position remained open and the employer continued to seek applicants from persons of complainant’s qualifications.

Id.

107. *Id.*; see also *Tex. Dep’t of Cmtv. Affairs v. Burdine*, 450 U.S. 248, 254-55 (1981) (explaining that the employer’s burden at this stage is simply a “burden of production” to introduce through admissible evidence that the plaintiff “was rejected, or someone else was preferred, for a legitimate, non[-]discriminatory reason.”).

108. *Burdine*, 450 U.S. at 256 (noting that the plaintiff “must have the opportunity to demonstrate that the proffered reason was not the true reason for the employment decision.”).

109. See *St. Mary’s Honor Ctr. v. Hicks*, 509 U.S. 502, 511 (1993) (explaining that the plaintiff’s *prima facie* case, along with his showing that the employer’s proffered non-discriminatory reason was pretextual, permits, but does not compel, the jury to “infer the ultimate fact of intentional discrimination”).

110. See, e.g., *Lindsey Watkins, Case Note, Employment Discrimination—Age Discrimination—The Fifth Circuit Holds a Plaintiff May Utilize the Mixed-Motives Method of Analysis in Age Discrimination Cases, Absent Any Direct Evidence of Discrimination: Rachid v. Jack in the Box, Inc.*, 58 SMU L. REV. 487, 487 (2005).

111. See *Filter Specialists, Inc. v. Brooks*, 906 N.E.2d 835, 839 (Ind. 2009) (noting that “[t]here are presently two alternative methods of establishing liability in a federal Title VII case. A plaintiff may pursue a ‘single motive’ theory of discrimination. Or a plaintiff may pursue a ‘mixed-motive’ theory of discrimination.” (internal citation omitted)).

112. 490 U.S. 228 (1989), superseded by statute as stated in *Landgraf v. U.S. Film Prods.*, 511 U.S. 244 (1994).

113. In explaining the “motivating factor” concept, the Court stated, “[W]e . . . know that Title VII meant to condemn even those decisions based on a mixture of legitimate and illegitimate considerations. When, therefore, an employer considers both . . . [illegitimate] and legitimate factors at the time of making a decision, that decision was ‘because of’ . . . [the illegitimate factor]

make a *prima facie* showing, and the defendant is required to articulate a legitimate non-discriminatory reason.¹¹⁴ But instead of showing that the defendant's proffered reason is pretextual, the plaintiff can show that "the defendant's reason, while true, is only *one* of the reasons for its conduct, and another 'motivating factor' is the plaintiff's protected characteristic."¹¹⁵ Under *Price Waterhouse*, if the plaintiff makes this motivating factor showing, the employer was afforded an affirmative defense to demonstrate that it would have made the same decision regardless of the plaintiff's protected trait.¹¹⁶ This differs from the pretext context, where the employer's burden to articulate a legitimate non-discriminatory reason is merely a burden of production; the employer in a mixed-motive case bears both the burdens of production and persuasion with respect to the affirmative defense.¹¹⁷

In 1991, Congress amended Title VII to codify the motivating factor standard.¹¹⁸ In the process, it eliminated the affirmative defense with respect to liability, making the defense relevant only to the issue of damages.¹¹⁹ Prior to *Gross*, courts had traditionally applied the full *Price Waterhouse* framework to ADEA age discrimination suits.¹²⁰ Congress did not amend the ADEA in 1991

and the other, legitimate considerations" *Id.* at 241.

114. See *Carey v. FedEx Ground Package Sys., Inc.*, 321 F. Supp. 2d 902, 916 (S.D. Ohio 2004) (explaining how mixed-motives comes into play in the "third stage . . . once the defendant has produced a legitimate, non[-]discriminatory reason for its conduct").

115. *Filter Specialists*, 906 N.E.2d at 842 (emphasis in original).

116. *Price Waterhouse*, 490 U.S. at 258 (holding that once the plaintiff makes the motivating factor showing, "the defendant may avoid a finding of liability only by proving by a preponderance of the evidence that it would have made the same decision even if it had not taken the plaintiff's . . . [protected trait] into account.").

117. See, e.g., *Weston-Smith v. Cooley Dickinson Hosp.*, Inc., 282 F.3d 60, 64 (1st Cir. 2002) (noting that under the mixed-motives method, once an employee makes the motivating factor showing, "the employer may then assert an affirmative defense, bearing the burdens of production and persuasion" (citations omitted)).

118. Civil Rights Act of 1991, Pub. L. No. 102-166, §107, 105 Stat. 1071, 1075 (codified as amended at 42 U.S.C. § 2000e-2(m) (2006)).

119. See 42 U.S.C. § 2000e-2(m) (stating that "an unlawful employment practice is established when the complaining party demonstrates that race, color, religion, sex, or national origin was a motivating factor for any employment practice, even though other factors also motivated the practice"); *id.* § 2000e-5(g)(2)(B). Section 2000e-5(g)(2)(B) states that

[o]n a claim in which an individual proves a violation under . . . [42 U.S.C. §] 2000e-2(m) . . . and a respondent demonstrates that the respondent would have taken the same action in the absence of the impermissible motivating factor, the court—

(i) may grant declaratory relief, injunctive relief (except as provided in clause (ii)), and attorney's fees and costs demonstrated to be directly attributable only to the pursuit of a claim under . . . [42 U.S.C. §] 2000e-2(m) . . . and

(ii) shall not award damages or issue an order requiring any admission, reinstatement, hiring, promotion, or payment

120. See Brief for the United States as Amicus Curiae Supporting Petitioner at 11, *Gross v.*

as it did Title VII, which explains why the courts continued to grant the full affirmative defense in ADEA cases.¹²¹ The issue in the circuits leading up to *Gross* was whether a plaintiff needed to present direct or merely circumstantial evidence of discrimination in order to utilize the mixed-motive framework.¹²²

B. Gross v. FBL Financial Services, Inc.

Gross presented the Supreme Court with the opportunity to settle the relatively modest direct versus circumstantial evidence issue, but the Supreme Court did not actually decide this issue. Jack Gross had worked for FBL Financial Services for thirty-two years (from 1971 to 2003) when, at the age of fifty-four, he was transferred from the position of claims administration director to claims project coordinator.¹²³ FBL transferred Gross’s job duties to a younger woman whom Gross had previously supervised.¹²⁴ Even though Gross’s pay was not reduced, he considered this reassignment a demotion and filed a claim for disparate treatment under the ADEA in the United States District Court for the Southern District of Iowa.¹²⁵

At the district court level, the court instructed the jury on the motivating factor standard, and the jury returned a verdict for Gross.¹²⁶ On appeal, the Eighth Circuit overruled this decision, finding that the trial court erroneously tendered the motivating factor instruction.¹²⁷ In the Eighth Circuit’s view, a motivating factor instruction is proper only where the plaintiff presents direct evidence “sufficient to support a finding by a reasonable fact finder that an

FBL Fin. Servs., Inc., 129 S. Ct. 2343 (2009) (No. 08-441), 2009 WL 253859 at *11 (noting that “[t]he lower courts . . . unanimously have agreed that the mixed-motive approach endorsed in *Price Waterhouse* applied to claims under the ADEA.”).

121. See *Baqir v. Principi*, 434 F.3d 733, 745 n.13 (4th Cir. 2006) (noting that “[b]ecause Congress did not similarly amend the ADEA [in 1991], however, ADEA mixed-motive cases remain subject to the *Price Waterhouse* analysis.” (citation omitted)).

122. In *Desert Place, Inc. v. Costa*, 539 U.S. 90, 101-02 (2003), a Title VII case, the Supreme Court held that direct evidence was not required in order to get a mixed-motive instruction. The circuits were split, however, as to whether *Desert Palace* applied to ADEA mixed-motive suits. See *King v. United States*, 553 F.3d 1156, 1160 (8th Cir. 2009) (applying mixed-motive to ADEA suits in cases where the plaintiff had direct evidence of discrimination); *Rachid v. Jack in the Box, Inc.*, 376 F.3d 305, 311 (5th Cir. 2004) (stating that “direct evidence of discrimination is not necessary to receive a mixed-motive analysis for an ADEA claim” (citation omitted)).

123. *Gross v. FBL Fin. Servs., Inc.*, 129 S. Ct. 2343, 2346 (2009).

124. *Id.* at 2346-47.

125. *Gross v. FBL Fin. Servs. Grp.*, No. 4:04-CV-60209-TJS, 2006 U.S. Dist. LEXIS 98081, at *1 (S.D. Iowa June 23, 2006), *rev’d sub nom. Gross v. FBL Fin. Servs., Inc.*, 526 F.3d 356 (8th Cir. 2008), *vacated*, 129 S. Ct. 2343 (2009).

126. *Id.* at *2.

127. *Gross v. FBL Fin. Servs., Inc.*, 526 F.3d 356, 362 (8th Cir. 2008), *vacated*, 129 S. Ct. 2343 (2009).

illegitimate criterion actually motivated' the adverse employment action.”¹²⁸ Therefore, according to the Eighth Circuit, the jury should only have been instructed as to whether age was the determining factor in FBL’s decision to demote Gross.¹²⁹

Gross appealed, and the Supreme Court granted certiorari.¹³⁰ On appeal, the parties framed the issue as “whether a plaintiff must ‘present direct evidence of discrimination in order to obtain a mixed-motive instruction in a non-Title VII case.’”¹³¹ The Supreme Court, however, did not decide this issue. Instead, the Court reframed the issue as whether the mixed-motive framework is even applicable to disparate treatment cases under the ADEA.¹³² Ultimately, the Supreme Court, in a five-to-four decision, held that it is not.¹³³

In coming to this conclusion, the Court employed both a textual analysis of the ADEA and an analysis of the legislative history of Title VII as it relates to the ADEA. With regard to the textual analysis, the Court noted that the relevant language in the ADEA—“because of”—means “by reason of: on account of.”¹³⁴ Thus, according to the Court, “the ordinary meaning of the ADEA’s requirement that an employer took adverse action ‘because of’ age is that age was the ‘reason’ that the employer decided to act.”¹³⁵ According to the Court, it follows that “[t]o establish a disparate-treatment claim under the plain language of the ADEA,

128. *Id.* at 359 (quoting *Thomas v. First Nat'l Bank of Wynne*, 111 F.3d 64, 66 (8th Cir. 1997)). What constitutes “direct evidence” of discriminatory intent is a somewhat murky concept in employment discrimination law because an employer will hardly ever come right out and say, “I am firing you because you are [insert protected trait].” See *Powe v. Ga. Pac. Co.*, 488 F. Supp. 467, 474 (W.D. Mich. 1980) (noting that intent must usually be proven indirectly through circumstantial evidence since “most discriminators before the courts today are too sophisticated to admit to discriminatory intent either on the witness stand or in their inner-company memoranda.”). That said, the Eighth Circuit defined direct evidence as evidence that “show[s] a specific link between the alleged discriminatory animus and the challenged decision.” *Gross*, 526 F.3d at 359 (quoting *Thomas*, 111 F.3d at 66).

129. *Gross*, 526 F.3d at 362.

130. *Gross v. FBL Fin. Servs., Inc.*, 129 S. Ct. 680 (2008).

131. *Gross v. FBL Fin. Servs., Inc.*, 129 S. Ct. 2343, 2348 (2009) (quoting petition for writ of certiorari).

132. *Id.* In justifying this decision, which is highly unusual for the Court, the Court stated, “Although the parties did not specifically frame the question to include this threshold inquiry, ‘[t]he statement of any question presented is deemed to comprise every subsidiary question fairly included therein.’” *Id.* at 2348 n.1 (quoting SUP. CT. R. 14.1). For an interesting take on how FBL convinced the Court to shift focus, see Michael Zimmer, *The Employer’s Strategy in Gross v. FBL Financials*, CONCURRING OPINIONS (Nov. 4, 2009, 10:43 AM), <http://www.concurringopinions.com/archives/2009/11/the-employers-strategy-in-gross-v-fbl-financials.html>.

133. *Gross*, 129 S. Ct. at 2352.

134. *Id.* at 2350 (citing 1 WEBSTER’S 3D NEW INT’L DICTIONARY 194 (1966)).

135. *Id.* (citing *Hazen Paper Co. v. Biggins*, 507 U.S. 604, 610 (1993) (explaining that for “because of” to be satisfied, the employee’s protected trait must have “had a determinative influence on the outcome”)).

therefore, a plaintiff must prove that age was the ‘but-for’ cause of the employer’s adverse decision.”¹³⁶

In so holding, the Court disregarded its prior decision in *Price Waterhouse*, which interpreted the exact same “because of” language in an almost identical setting.¹³⁷ The *Gross* Court found it instructive that the *Price Waterhouse* Court interpreted the “because of” language under Title VII, whereas the language at issue here was under the ADEA.¹³⁸ If this analysis were to leave one unsatisfied and wondering why the exact same language in two nearly identical statutes should be interpreted so differently, the Court addressed that issue when it stated:

[W]e reject petitioner’s contention that our interpretation is controlled by *Price-Waterhouse*, which initially established that the burden of persuasion shifted in alleged mixed-motive Title VII claims. In any event, it is far from clear that the Court would have the same approach were it to consider the question today in the first instance.¹³⁹

Thus, not only was *Price Waterhouse* inapplicable because it was a Title VII case, its interpretation of the “because of” language was also probably wrong, according to the *Gross* Court.

The majority in *Gross* was also persuaded by the fact that when Congress codified *Price Waterhouse*’s mixed-motive motivating factor standard in the 1991 Civil Rights Act, it failed to similarly amend the ADEA. According to the majority, this failure to similarly amend was affirmative evidence that Congress did not intend for the motivating factor to extend to the ADEA because “[w]hen Congress amends one statutory provision but not another, it is presumed to have acted intentionally.”¹⁴⁰ This principle is especially true, according to the majority, when two statutes are amended simultaneously (as the ADEA and Title VII were in 1991) because “‘negative implications raised by disparate provisions are strongest’ where the provisions were ‘considered simultaneously when the language raising the implications was inserted.’”¹⁴¹ Thus, the fact that Congress neglected to amend the ADEA to reflect the motivating factor standard in 1991 was the second major reason the majority found to deny mixed-motive claims under the ADEA. The Court stated its holding as follows:

We hold that a plaintiff bringing a disparate treatment claim pursuant to the ADEA must prove, by a preponderance of the evidence, that age was the “but-for” cause of the challenged adverse employment action. The

136. *Id.* (citing *Bridge v. Phx. Bond & Indem. Co.*, 553 U.S. 639, 652 (2008) (noting that the words “by reason of” require at least a showing of “but-for” causation)).

137. See *Price Waterhouse v. Hopkins*, 490 U.S. 228, 231-37 (1989), superseded by statute as stated in *Landgraf v. U.S. Film Prods.*, 511 U.S. 244 (1994).

138. See *Gross*, 129 S. Ct. at 2349 (noting that “this Court’s interpretation of the ADEA is not governed by Title VII decisions such as *Price Waterhouse*”).

139. *Id.* at 2351-52 (citations omitted).

140. *Id.* at 2349.

141. *Id.* at 2345 (quoting *Lindh v. Murphy*, 521 U.S. 320, 330 (1997)).

burden of persuasion does not shift to the employer to show that it would have taken the action regardless of age, even when a plaintiff has produced some evidence that age was one motivating factor in that decision.¹⁴²

Simply put, the *Gross* Court determined that the *Price Waterhouse* mixed-motive framework no longer applies to ADEA disparate treatment claims.

C. Implications of Gross

The *Gross* decision does not bode well for age discrimination plaintiffs who file their claims under the ADEA—which almost all Indiana employees are forced to do so by the IADA’s provision exempting employers “subject to” the ADEA from its coverage—because the “but-for” standard requires a higher showing of intent than the motivating factor standard. According to one commentator, after *Gross*, “[t]he employee will have to prove that age discrimination, rather than cost savings, or efficiency or something [else], was not only a cause, not only the significant cause, not only the motivating cause, but the exclusive cause of an adverse employment action.”¹⁴³ Ultimately, *Gross* will make it “more difficult for individuals to bring successful workplace age discrimination cases against their employers.”¹⁴⁴ Indiana employees are especially affected by the *Gross* decision because they have no alternative state statute under which they can pursue their claims. Indiana is unique in this regard, as the following section demonstrates.

III. HOW INDIANA’S NEIGHBORING STATES TREAT AGE DISCRIMINATION

Every state in the United States has a statute that prohibits age discrimination.¹⁴⁵ If Indiana recognizes the deficiencies in its age discrimination regime and decides to take steps to expand protection, it can learn from how other states deal with the problem. This section briefly highlights the pertinent anti-age discrimination statutes and cases that interpret those statutes (Indiana’s Midwest neighbors—Kentucky, Ohio, Michigan, and Illinois) to briefly show how Indiana’s age discrimination protections are inferior to other states’ measures.

A. Kentucky

Kentucky prohibits age discrimination along with the other major types of

142. *Id.* at 2352.

143. Kimberly Atkins, *Surprise Ruling Makes Age Bias Cases Tough for Plaintiffs*, MINN. LAW., June 29, 2009; *see also* Campolongo, *supra* note 9.

144. Steven D. Irwin, *ADEA Decision Ill-Timed in Tougher Economy*, P.A. LAW WKLY., Aug. 24, 2009, at A13.

145. For a comprehensive listing of all fifty states’ age discrimination statutes, see Hillina Tadesse Tamrat, Note, *Sovereign Immunity Under the Eleventh Amendment: Kimel and Garrett, What Next for State Employees?*, 11 ELDER L.J. 171, 183 n.126 (2003).

discrimination in the Kentucky Civil Rights Act (KCRA).¹⁴⁶ The KCRA was modeled after Title VII and provides:

It is an unlawful employment practice for an employer . . . [t]o fail or refuse to hire, or to discharge any individual, or otherwise to discriminate against an individual with respect to compensation, terms, conditions, or privileges of employment, because of the individual's race, color, religion, national origin, sex, *age forty (40) and over*, because the person is a qualified individual with a disability, or because the person is a smoker or nonsmoker. . . .¹⁴⁷

If an individual in Kentucky believes he has been the victim of age discrimination, the KCRA affords the individual administrative remedies or a private right of action, whichever the complainant prefers.¹⁴⁸ In the event that age discrimination is found to have occurred, the KCRA allows the plaintiff to recover “actual damages sustained, together with the costs of the law suit,” as well as attorney fees.¹⁴⁹ The Kentucky courts have interpreted “actual damages” to mean full compensatory damages, including damages for mental and emotional distress.¹⁵⁰ Punitive damages, however, are not available.¹⁵¹

B. Ohio

Ohio prohibits age discrimination under two separate statutory provisions within its civil rights act. First, age discrimination is included with other forms of discrimination in section 4112.02(A) of Ohio’s code, which provides:

It shall be an unlawful discriminatory practice . . . [f]or any employer, because of the race, color, religion, sex, military status, national origin, disability, *age*, or ancestry of any person, to discharge without just cause, to refuse to hire, or otherwise discriminate against that person with respect to hire, tenure, terms, conditions, or privileges of employment, or any matter directly or indirectly related to employment.¹⁵²

Age discrimination is also singled out for special treatment in section 4112.14(A) of Ohio’s code, which provides:

146. KY. REV. STAT. ANN. §§ 344.010 to -.500 (West, Westlaw through 2010 legislation).

147. *Id.* § 344.040(1)(a) (emphasis added).

148. See *id.* § 344.450; see also *Canamore v. Tube Turns Div. of Chemetron Corp.*, 676 S.W.2d 800, 804 (Ky. Ct. App. 1984) (noting that section 344.450 “is intended to give those individuals who do not wish to proceed before the . . . [Kentucky Human Rights Commission] an opportunity in circuit court to have the fullest range of remedies allowable”).

149. KY. REV. STAT. ANN § 344.450.

150. See *Childers Oil Co. v. Adkins*, 256 S.W.3d 19, 28 (Ky. 2008) (utilizing the *McDonnell Douglas* framework regarding age discrimination).

151. *Id.* at 26-27; *Ky. Dep’t of Corr. v. McCullough*, 123 S.W.3d 130, 140 (Ky. 2003).

152. OHIO REV. CODE ANN. § 4112.02(A) (West, Westlaw through 2010 legislation) (emphasis added).

No employer shall discriminate in any job opening against any applicant or discharge without just cause any employee aged forty or older who is physically capable to perform the duties and otherwise meets the established requirements of the job and laws pertaining to the relationship between employer and employee.¹⁵³

Both sections provide an age complainant with a private right of action to pursue his claim in court.¹⁵⁴ A complainant may not, however, institute a civil action under section 4112.14 if he has the opportunity to arbitrate the claim or if the dispute has already been arbitrated.¹⁵⁵ In addition, if an age complainant so chooses, he may elect to pursue administrative remedies with the Ohio Civil Rights Commission.¹⁵⁶

Successful discrimination plaintiffs in Ohio, including age discrimination plaintiffs, are “entitled to ‘make whole’ relief.”¹⁵⁷ Therefore, in addition to back pay, a successful age discrimination plaintiff in Ohio is also entitled to reinstatement or front pay in the event that reinstatement is inappropriate.¹⁵⁸ Attorney fees are not explicitly provided for by statute as required by the “American rule” on attorney fees,¹⁵⁹ which Ohio follows,¹⁶⁰ but Ohio makes an exception to the rule in civil rights cases where punitive damages are proper.¹⁶¹ Punitive damages are available upon a showing of actual malice.¹⁶²

C. Michigan

Michigan bans age discrimination in its primary anti-discrimination statute, the Elliott-Larsen Civil Rights Act.¹⁶³ Section 202(1) of the act states that

[a]n employer shall not do any of the following: (a) Fail or refuse to hire or recruit, discharge, or otherwise discriminate against an individual with respect to employment, compensation, or a term, condition, or privilege

153. *Id.* § 4112.14(A).

154. *Id.* § 4112.02(N); *id.* § 4112.14(B).

155. *Id.* § 4112.14(C).

156. *Id.* § 4112.05(B)(1).

157. *Potocnik v. Sifco Indus., Inc.*, 660 N.E.2d 510, 518 (Ohio Ct. App. 1995).

158. *See id.*

159. The so-called “American rule” on attorney fees states that “the prevailing party cannot recover attorney fees as part of the cost of litigation in the absence of statutory authorization.” *Sutherland v. Nationwide Gen. Ins. Co.*, 657 N.E.2d 281, 282 (Ohio Ct. App. 1995) (citing *Sorin v. Bd. of Educ. of Warrensville Heights Sch. Dist.*, 347 N.E.2d 527, 528-29 (Ohio 1976)).

160. *See id.* at 283 (noting that “Ohio courts have held fast to the ‘American rule’ proscribing an award of attorney fees in the absence of statutory authorization”).

161. *Id.*

162. *Srail v. RJF Int'l Corp.*, 711 N.E.2d 264, 274 (Ohio Ct. App. 1998) (age discrimination case).

163. MICH. COMP. LAWS ANN. §§ 37.2102 to -.2804 (West, Westlaw through 2010 legislation).

of employment, because of religion, race, color, national origin, *age*, sex, height, weight, or marital status.¹⁶⁴

The Michigan act provides for a private cause of action.¹⁶⁵ Successful plaintiffs can recover damages for any “injury or loss caused by each violation of . . . [the Elliott-Larsen Civil Rights Act], including reasonable attorney’s fees.”¹⁶⁶ Those damages may include “damages for humiliation, embarrassment, outrage and disappointment as well as loss of wages, loss of pension rights and employee benefits, loss of seniority and loss of employment.”¹⁶⁷ In addition, costs, including attorney and witness fees, are recoverable.¹⁶⁸ Punitive damages, however, are not recoverable.¹⁶⁹

D. Illinois

Age discrimination in Illinois is addressed alongside other forms of discrimination in the Illinois Human Rights Act (IHRA).¹⁷⁰ The IHRA prohibits “unlawful discrimination,”¹⁷¹ which includes discrimination based on “race, color, religion, sex, national origin, ancestry, *age*, order of protection status, marital status, physical or mental disability, military status, sexual orientation, or unfavorable discharge from military service in connection with employment, real estate transactions, access to financial credit, and the availability of public accommodations.”¹⁷²

An individual alleging age discrimination in Illinois can choose to file a charge with the Illinois Human Rights Commission or institute a private civil action.¹⁷³ If a complainant files a charge with the commission and the commission finds that discrimination has occurred, the commission is authorized to enter a cease and desist order, award actual damages as well as attorney fees

164. *Id.* § 37.2202(1)(a) (emphasis added).

165. Section 801 states, “A person alleging a violation of this act may bring a civil action for appropriate injunctive relief or damages, or both.” *Id.* § 37.2801(1).

166. *Id.* § 37.2801(3).

167. Schafke v. Chrysler Corp., 383 N.W.2d 141, 143 (Mich. Ct. App. 1985) (age discrimination case).

168. MICH. COMP. LAWS ANN. § 37.2802.

169. See Gilbert v. Daimler Chrysler Corp., 685 N.W.2d 391, 400 (Mich. 2004) (explaining that in Michigan, punitive damages are only recoverable when expressly authorized by statute, and “the Civil Rights Act does not authorize punitive damages”).

170. 775 ILL. COMP. STAT. ANN. 5/1-101 to 5/10-104 (West, Westlaw through 2010 legislation).

171. *Id.* 5/2-102(A).

172. *Id.* 5/1-102(A) (emphasis added).

173. *Id.* 5/10-102(A)(1) (“An aggrieved party may commence a civil action in an appropriate [c]ircuit [c]ourt not later than 2 years after the occurrence or the termination of an alleged civil rights violation. . . .”); *id.* 5/7B-102(A)(1) (“Within one year after the date that a civil rights violation allegedly has been committed or terminated, a charge in writing under oath or affirmation may be filed with the [d]epartment by an aggrieved party. . . .”).

and costs, and assess a civil penalty.¹⁷⁴ Actual damages include damages for emotional harm and mental suffering.¹⁷⁵ If, on the other hand, a complainant files a civil action and a court finds a violation, the court is authorized to award actual and punitive damages, injunctive relief, and attorney fees and costs.¹⁷⁶

IV. AGE DISCRIMINATION SHOULD BE ADDRESSED

A. *The Problem*

As this Note has thus far demonstrated, Indiana treats age discrimination far less seriously than the types of discrimination covered by the ICRL, and it lags far behind other states in terms of the protection it affords older workers. One could argue, however, that age discrimination is materially different from other forms of discrimination in that discrimination based on race, sex, religion, and national origin is prejudice-based, whereas age discrimination is generally based on unfounded stereotypes.¹⁷⁷ It follows, then, that age discrimination should be kept separate from other types of discrimination. Professor Rhonda Reaves, echoing this sentiment, argues against what she calls “cross-contamination,” which is “[t]he failure to recognize relevant differences” between age and race discrimination, and which “can have the unintended effect of undermining fundamental principles of anti-discrimination law.”¹⁷⁸ However, even Reaves acknowledges that age discrimination laws are necessary;¹⁷⁹ she just urges the courts to acknowledge the fundamental differences between age and race discrimination and not to allow flexible principles in the age context to seep into the race context.¹⁸⁰

The argument against mixing stereotype-based discrimination with prejudice-based discrimination is reasonable. However, this does not mean that qualified older workers are any less deserving of protection.¹⁸¹ As one writer points out,

174. *Id.* 5/8B-104(A)-(D).

175. See *Szkoda v. Ill. Human Rights Comm'n*, 706 N.E.2d 962, 972 (Ill. App. Ct. 1998) (citing *Arlington Park Race Track Corp. v. Human Rights Comm'n*, 557 N.E.2d 517, 524 (Ill. Ct. App. 1990)) (“The ‘actual damages’ provision of section 8B-104(B) of the Act also includes damages for emotional harm and mental suffering.”)).

176. 775 ILL. COMP. STAT. ANN. 5/10-102(C)(1)-(2).

177. See REPORT OF THE SEC'Y OF LABOR, *supra* note 99, at 2 (finding that age discrimination occurs “because of assumptions about the effect of age on their ability to do a job *when there is in fact no basis for these assumptions*,” not “prejudice based on dislike or intolerance of the older worker”).

178. Rhonda M. Reaves, *One of These Things Is Not Like the Other: Analogizing Ageism to Racism in Employment Discrimination Cases*, 38 U. RICH. L. REV. 839, 844 (2004).

179. *Id.* at 902.

180. *Id.*

181. A few words of caution are in order here. I am merely advocating the modest proposal that qualified older workers should not have adverse employment actions taken against them “because of” their age. As with all discrimination laws, this does not mean that an employer would

“evidence of stereotypical thinking supports an ultimate inference of intent to discriminate precisely because it is an unconscious expression of bias.”¹⁸² Indeed, Congress enacted the ADEA on the assumption that age discrimination was the result of stereotypes (not prejudice), yet Title VII and the ADEA share a common purpose: “the elimination of discrimination in the workplace.”¹⁸³ In any case, the effects of age discrimination are more important than the cause. And the consequences of age discrimination are real, as evidenced by the EEOC’s statistics.¹⁸⁴ Moreover, in 2009, the number of older workers who elected to collect Social Security benefits early increased by nineteen percent.¹⁸⁵ As Richard Johnson, senior fellow at the Urban Institute, commented, “[t]here are just not enough jobs for older people. They have no choice but to go on Social Security.”¹⁸⁶ The economy does not fully explain this, as there are studies that show that employers are less likely to hire older workers than younger workers.¹⁸⁷ The problem of age discrimination is only exacerbated by the fact that age fifty-five and over is the fastest growing age segment in the American workforce.¹⁸⁸

Professor Joanna Lahey summed up the problem well when she commented that “employers clearly do *treat older workers differently*[,] and the impact can be really harmful, especially for those with low savings who most need work.”¹⁸⁹ Indeed, “[t]he essence of discrimination is that otherwise similarly situated

be forced to hire, promote, or retain older workers who are unqualified. It just means that an employer cannot discriminate on the basis of a worker’s age. See *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973) (requiring a discrimination plaintiff to show he was qualified as an element of the *prima facie* case); *Sommer v. City of Elkhart*, No. 3:08-CV-522, 2009 WL 5200525, at *4 (N.D. Ind. Dec. 23, 2009) (noting that “[i]f it is true that . . . [the plaintiff] was unqualified and that she failed to even apply for the positions she claims she was denied, then her claim fails.”).

182. Michael J. Zimmer, *A Chain of Inferences Proving Discrimination*, 79 U. COLO. L. REV. 1243, 1279 (2008).

183. *McKennon v. Nashville Banner Publ’g Co.*, 513 U.S. 352, 358 (1995) (quoting *Oscar Mayer & Co. v. Evans*, 441 U.S. 750, 756 (1979)).

184. EEOC, CHARGE STATISTICS, *supra* note 1; *see also supra* notes 1-3.

185. Richard Wolf, *Social Security Collectors Up 19%*, U.S.A. TODAY, Oct. 1, 2009, at A1, available at http://www.usatoday.com/news/nation/2009-10-01-social-security_N.htm.

186. *Id.*

187. *See, e.g., JOANNA N. LAHEY, CTR. FOR RETIREMENT RESEARCH AT BOS. COLL., DO OLDER WORKERS FACE DISCRIMINATION?* 3 (July 2005), available at http://crr.bc.edu/images/stories/Briefs/ib_33.pdf. Lahey conducted a study in which she sent out four thousand resumes to employers in Boston, Massachusetts and St. Petersburg, Florida. *Id.* The age of the applicants on the resumes, indicated by their date of high school graduation, ranged from thirty-five to sixty-two. *Id.* Lahey found that, all else equal, a worker over the age of fifty was forty percent less likely to be called back. *See id.*

188. *See U.S. DEP’T OF LABOR, EMPLOYMENT PROJECTIONS: CIVILIAN LABOR FORCE BY AGE, SEX, RACE, AND ETHNICITY*, <http://www.bls.gov/emp/emplab06.htm> (last modified Dec. 8, 2010).

189. LAHEY, *supra* note 187, at 4 (emphasis added).

individuals are treated differently because of their race, sex, religion, or age.”¹⁹⁰ In the end, whether age discrimination is prejudice-based or stereotype-based should be largely irrelevant to the issue of whether it should be tolerated. With such anemic age discrimination laws, especially with regard to state employees, Indiana effectively tolerates age discrimination. Indiana has no good reason to have such weak age discrimination laws.

B. *The Solutions*

Indiana has two fundamental deficiencies within its age discrimination regime. First, state employees are practically left in the cold as a result of the IADA provision that exempts employers “subject to” the ADEA because of *Kimel*’s holding that sovereign immunity bars ADEA actions by state employees against their employers in federal court. Second, the IADA does not provide the option for a private cause of action or meaningful remedies even for the employees who are covered by it, and age is not a protected class under the ICRL. Indiana should adopt legislation to rectify these shortcomings. If the Indiana General Assembly fails to do so, the courts should revisit the issue of whether termination based on age could form the basis of a common law wrongful discharge claim. This section explores both options.

1. *Legislative Solutions.*—The most ideal option for Indiana would be to adopt legislation to rectify the IADA’s shortcomings. The easiest and most effective action would be to simply add age as a protected class under the ICRL, as Kentucky, Ohio, Michigan, and Illinois do in their respective civil rights statutes. This addition would provide relief for employees unable to proceed under the federal ADEA—either because of sovereign immunity or because their employers utilize fewer than twenty employees—because under the ICRL, “[e]mployer” means the state or any political or civil subdivision thereof and any person employing six (6) or more persons within the state”¹⁹¹ This approach would also make available the same remedies to age discrimination victims as are currently available to race, color, religion, sex, disability, national origin, and ancestry plaintiffs under the ICRL. Such remedies would include reinstatement and recovery of lost wages, salaries, and commissions.¹⁹² However, this approach would not allow prevailing plaintiffs to recover attorney fees (as the ICRL currently does not have an attorney fee provision), non-economic damages, or punitive damages.¹⁹³ Despite the aforementioned shortcomings of the ICRL,¹⁹⁴ adding age to the list of categories covered by the ICRL would best protect age discrimination victims because it would put age discrimination on par with race,

190. *Cooper v. Oak Park Sch. Dist.*, 624 F. Supp. 515, 516 (E.D. Mich. 1986) (citing *C. Thorrez Indus., Inc. v. Mich. Dep’t of Civil Rights*, 278 N.W.2d 725, 727 (Mich. Ct. App. 1979)).

191. IND. CODE § 22-9-1-3(h) (2010).

192. *Id.* § 22-9-1-6(K)(A); *see also* *Ind. Civil Rights Comm’n v. Midwest Steel Div.*, 450 N.E.2d 130, 140-41 (Ind. Ct. App. 1983).

193. *See supra* Part I.A-D.

194. *See generally* Olivier, *supra* note 90.

color, sex, national origin, disability, national origin, and ancestry discrimination. This action would thereby eliminate the anomaly that is age discrimination under current Indiana law.

Indeed, in 2009, a bipartisan bill that would have made age a protected category under the ICRL was introduced and considered in the Indiana General Assembly.¹⁹⁵ The bill’s authors noted the increasing problem of age discrimination, and they argued that it was unfair to have such a discrepancy in the way age discrimination is treated under the IADA compared to other forms under the ICRL.¹⁹⁶ This legislation passed the Indiana House of Representatives by a vote of fifty-eight to thirty-seven,¹⁹⁷ but senate amendments severely weakened it, stripping the provision adding age as a protected category under the ICRL.¹⁹⁸ In the end, the bill only increased the maximum age limit for age discrimination claims from seventy to seventy-five and added a provision allowing the Indiana Commissioner of Labor to publicize the results of age discrimination complaints.¹⁹⁹ Thus, it appears that the most effective method for strengthening Indiana’s protections against age discrimination is probably unrealistic in Indiana’s current political climate.²⁰⁰

There are, however, other legislative measures that might be able to pass the Indiana General Assembly. For instance, the legislature could amend the IADA to rectify the statute’s current shortcomings. In order to do this, the remedial authority of the Indiana Commissioner of Labor must be enhanced. As it stands

195. H.B. 1014, 116th Gen. Assemb., Reg. Sess. (Ind. 2009). This bill was authored by Representatives Vernon Smith (D-Gary), David Niezgodski (D-South Bend), and Ed Soliday (R-Valparaiso).

196. See Press Release, Indiana House Democrats, Smith Fighting to Protect Hoosiers Against Age Discrimination (Feb. 3, 2009) (on file with author).

197. Ind. H. Roll Call, H.B. 1014, 116th Gen. Assemb., Reg. Sess. (Ind. 2009).

198. Opponents of the bill argued that it would cost employers too much money to defend age discrimination complaints and that some workers might use the new law as “a weapon to extort money from the people not equipped to handle these claims.” Richard Gootee, *Bill Aims to Bulk Up Age-Bias Laws*, INDY.COM (Apr. 2, 2009), <http://www.indy.com/posts/bill-aims-to-bulk-up-age-bias-laws>. Ironically, opponents of the bill also argued that it was unnecessary because of the “options already in place . . . including federal anti-discrimination laws.” *Id.*

199. H.B. 1014, 116th Gen. Assemb., Reg. Sess. (Ind. 2009) (as passed by Indiana House of Representatives, Apr. 29, 2009).

200. The Indiana General Assembly did pass one bill targeted at age discrimination in the 2010 session, however. House Bill 1005, authored by Representative Vernon Smith, removes a provision from Indiana law that eliminated a teacher’s indefinite contract once the teacher turned seventy-one years old. H.B. 1005, 116th Gen. Assemb., 2d Reg. Sess. (Ind. 2010) (as passed by Indiana House of Representatives, Jan. 26, 2010). This provision discriminated on the basis of age because it forced older teachers to work under one-year contracts for no other reason besides their age. See Press Release, Indiana House Democrats, House Passes Smith Bill to Eliminate Teacher Age Discrimination (Jan. 26, 2010), available at http://www.in.gov/legislative/house_democrats/smith_news_20100126.pdf. The bill was signed into state law in 2010. IND. CODE § 20-28-6-8 (2010).

now, the IADA “offers no remedy to an individual.”²⁰¹ This is unacceptable. At the very least, the Indiana Commissioner of Labor should be able to order reinstatement and damages for back pay. Alternatively, keeping in mind that the commissioner has responsibilities beyond mediating age discrimination disputes, the IADA could provide for a private right of action in state court with remedies limited to what is available under the ICRL. However, to do this, the IADA would have to provide for attorney fees—unlike the ICRL, which currently does not. Without a provision for attorney fees, alleged victims of age discrimination will be discouraged, if not precluded, from pursuing their claims in court. Indeed, Indiana’s neighboring states—Kentucky, Ohio, Michigan, and Illinois—all provide for attorney fees in their respective civil rights statutes.²⁰²

Finally, to blunt the impact of *Gross*, any age discrimination legislation that Indiana adopts should reflect the “motivating factor” standard of causation. In fact, Indiana courts currently recognize the mixed-motive framework in race, religion, color, sex, disability, national origin, and ancestry suits under the ICRL.²⁰³ Reflecting the “motivating factor” standard could easily be done by adopting language parallel to Title VII’s statutory motivating factor standard, which states that “an unlawful employment practice is established when the complaining party demonstrates that race, color, religion, sex, or national origin was a motivating factor for any employment practice, even though other factors also motivated the practice.”²⁰⁴ Legislation is pending in the United States Congress to codify the motivating factor standard in the ADEA,²⁰⁵ which, if passed, would moot the issue with respect to Indiana plaintiffs who can proceed under the ADEA. However, the “motivating factor” standard should still be included in any new legislation to afford equal treatment to those who are relegated to the Indiana statutes.

2. *Judicial Solutions.*—In *Montgomery*, the plaintiff argued that the IADA creates a private cause of action for termination in violation of public policy under *Frampton*.²⁰⁶ The Indiana Supreme Court rejected this argument for various reasons,²⁰⁷ chief among them that “[g]eneral expressions of public policy

201. *Hague v. Thompson Distrib. Co.*, No. 1:02-CV-01744-RLY-TA, 2005 WL 995689, at *7 (S.D. Ind. Feb. 9, 2005).

202. 775 ILL. COMP. STAT. ANN. 5/8B-104(D) (West, Westlaw through 2010 legislation); KY. REV. STAT. ANN. § 344.450 (West, Westlaw through 2010 legislation); MICH. COMP. LAWS ANN. § 37.2802 (West, Westlaw through 2010 legislation); *Sutherland v. Nationwide Gen. Ins. Co.*, 657 N.E.2d 281, 283 (Ohio Ct. App. 1995).

203. See *Filter Specialists, Inc. v. Brooks*, 906 N.E.2d 835, 839 (Ind. 2009) (noting that “a plaintiff may pursue a ‘mixed-motive’ theory of discrimination” and upholding the civil rights commission’s finding that the plaintiff’s race was the “motivating factor” behind the termination.).

204. 42 U.S.C. § 2000e-2(m) (2006).

205. See *Protecting Older Workers Against Age Discrimination Act*, H.R. 3721, 111th Cong. (2009). At the time of publication, this legislation remains in the phase of subcommittee hearings.

206. *Montgomery v. Bd. of Trs. of Purdue Univ.*, 849 N.E.2d 1120, 1127 (Ind. 2006).

207. See *supra* Part I.B.

do not support new exceptions to the employment-at-will doctrine.”²⁰⁸ On its face, this opinion might appear to foreclose any possibility that Indiana courts will recognize a wrongful discharge claim for age discrimination. However, in future cases, plaintiffs could argue that the IADA creates a right not to be fired because of age. Indeed, Indiana courts recognize a cause of action for wrongful discharge for employees who are terminated solely for exercising a statutorily conferred right.²⁰⁹ The *Montgomery* court did not have the opportunity to expressly reject this argument because Montgomery did not raise it in this way, instead arguing that his discharge violated “public policy.”²¹⁰

Other persuasive sources of law support the argument that the IADA creates a right to be free from age discrimination in employment and that therefore, there must be remedies to protect that right. In *Sullivan v. Little Hunting Park, Inc.*,²¹¹ the United States Supreme Court noted that “[t]he existence of a statutory right implies the existence of all necessary and appropriate remedies.”²¹² Another United States Supreme Court case, *Texas & Pacific Railway Co. v. Rigsby*,²¹³ elaborates on this concept: “A disregard of the command of the statute is a wrongful act, and where it results in damage to one of the . . . [classes] for whose especial benefit the statute was enacted, the right to recover the damages from the party in default is implied.”²¹⁴ In addition, the Restatement (Second) of Torts also supports creating a wrongful discharge cause of action for age discrimination:

When a legislative provision protects a class of persons by proscribing or requiring certain conduct but does not provide a civil remedy for the violation, the court may, if it determines that the remedy is appropriate in furtherance of the purpose of the legislation and needed to assure the effectiveness of the provision, accord to an injured member of the class a right of action, using a suitable existing tort action or a new cause of action analogous to an existing tort action.²¹⁵

Thus, Indiana’s own precedent, along with persuasive sources from the United States Supreme Court and the Restatement (Second) of Torts, supports the argument that Indiana should recognize a cause of action for wrongful discharge for termination based on age.

In addition to all of the above sources, the state of Ohio’s reasoning is particularly persuasive. In *Leininger v. Pioneer National Latex*,²¹⁶ a sixty-year-

208. *Montgomery*, 849 N.E.2d at 1128.

209. *Frampton v. Cent. Ind. Gas Co.*, 297 N.E.2d 425, 428 (Ind. 1973).

210. *Montgomery*, 849 N.E.2d at 1127.

211. 396 U.S. 229 (1969).

212. *Id.* at 239.

213. 241 U.S. 33 (1916).

214. *Id.* at 39; see also *Allen v. State Bd. of Elections*, 393 U.S. 544, 557 (1987) (noting that the Voting Rights Act “might well prove an empty promise unless the private citizen were allowed to seek judicial enforcement of the prohibition.”).

215. RESTATEMENT (SECOND) OF TORTS § 874A (1979).

216. 875 N.E.2d 36 (Ohio 2007).

old woman was fired from her human resources position after nineteen years with her employer and was then replaced by a twenty-one-year-old.²¹⁷ Consistent with Ohio precedent,²¹⁸ the plaintiff filed a common law wrongful discharge claim against the employer for age discrimination.²¹⁹ Despite earlier precedent in Ohio supporting an age discrimination cause of action for wrongful discharge, the Ohio Supreme Court held that a wrongful discharge cause of action for age discrimination no longer existed “because the remedies in . . . [the Ohio Civil Rights Act] provide complete relief for a statutory claim for age discrimination.”²²⁰ In so holding, the court reasoned that “[n]ow, remedies are available to plaintiffs such as Leininger, pursuant to . . . [the Ohio Civil Rights Act], that were not available to . . . [plaintiffs before Leininger].”²²¹ In other words, the reason there was no wrongful discharge cause of action was because the Ohio Civil Rights Act had been revised to provide relief for victims of age discrimination. The wrongful discharge cause of action was no longer necessary because doing away with it would not jeopardize public policy²²²—age discrimination was now adequately addressed by statute. In Indiana, on the other hand, age discrimination is currently not adequately protected by statute.²²³ Therefore, Indiana courts should recognize a wrongful discharge age discrimination cause of action to protect employees’ right to be free from age discrimination and Indiana’s public policy against age discrimination.

CONCLUSION

Withholding employment opportunities from someone because of his age, rather than his qualifications, is fundamentally wrong. Even conceding that age discrimination originates from unfounded stereotypes (rather than from prejudice, like some other forms of discrimination), age discrimination is still unacceptable and has adverse consequences to the victim and the overall economy. Although many Indiana workers are protected from age discrimination by the federal ADEA (albeit less so after *Gross*), many are not so protected as a result of sovereign immunity and the amount of employees their employers have. For those Indiana employees who are unable to proceed in federal court, they literally have no recourse under Indiana law. As Justice Rucker so fittingly put it, these Indiana employees find themselves in a “Catch-22.”²²⁴ Indiana should follow the

217. *Id.* at 38.

218. See *Gessner v. City of Union*, 823 N.E.2d 1 (Ohio 2004); *Livingston v. Hillside Rehab. Hosp.*, 680 N.E.2d 1220 (Ohio 1997), superseded by statute as stated in *Leininger*, 875 N.E.2d at 44.

219. *Leininger*, 875 N.E.2d at 38.

220. *Id.* at 44.

221. *Id.* at 40.

222. See *id.* at 39.

223. See *supra* Parts I.A, I.D.

224. *Montgomery v. Bd. of Trs. of Purdue Univ.*, 849 N.E.2d 1120, 1131 (Ind. 2006) (Rucker, J., dissenting). According to Justice Rucker, “[a] Catch-22 is generally understood as a no-win

example of its neighboring states and add age as a protected class under the ICRL. Alternatively, since that outcome appears to be unlikely as a political matter, Indiana should at least enhance the investigative and remedial authority of the Indiana Commissioner of Labor or provide an alternative route for age discrimination complainants to have their day in court. If the Indiana General Assembly does not take action, Indiana courts should protect both employees' right to be free from arbitrary age discrimination and Indiana's public policy against discrimination. Furthermore, they should allow age discrimination to form the basis of a common law wrongful discharge action.





